

-899-

Identities = 147/531 (27%), Positives = 261/531 (48%), Gaps = 26/531 (4%)

Query: 19 IRICKLTMSIRRVRYKEQYLFKQEEADKRIEFIEECSNTKGLAGKLRLLPQKVWLE 78
 I + K K+I++ R ++Y+++ + + IE+IE+ T G K++L QK W E
 5 Sbjct: 16 IELNKYMRKTIQKQIRIHKKYIYRYDRVTQAIEWIEDNFYLTIGNLMKIKLHPTQKYWYE 75

Query: 79 TTWGFYHTVEVTKTNPDTLEEYTDYEERRLIHEVPIIVPRGTGKTLTGSIAEVEGQIIDG 138
 G+ D ++E + LI+E+ + + RG+GK++L + I+ G
 10 Sbjct: 76 LMLGY-----DMVDEKG--VQVNLINELFLNLGRSGSKSLMATRVLNMILGG 122

Query: 139 EWGADIQLLAYSREQAGYLFNASRAMLSNEESLLHYMREADILRSTKQGIYETTNLSMS 198
 ++G + ++AY QA ++F+ R ++L Y E I +STKQG+ + +
 Sbjct: 123 QYGGESLVIAVDNTQARHVFDQVRNQTEASDTRLRVY-NENKIFKSTKQGLETAFAKTTFFK 181

Query: 199 IKTSDBESLDGTNAHYNIFDEVHTYDDDFIKVVDGSSSRKRKNWITWYISTNGTKRDKLF 258
 +T+D G N+ NIFDEVHTY +D + VN GS +K+ NW + YI++ G KR D L+
 15 Sbjct: 182 KQINDTLRAQGGNSSLNIFDEVHTYGEDITESVKNKGSRQKQDNWQSIYITSGGLKRDGLY 241

Query: 259 DKYYNIWVDILDDKIINDSVMPWIYQLDDVSEIHDPDMWQKAMPLLGITTEKETIARDIE 318
 DK + +++ ND +Y L++ ++ D W A+PL+G + + + E
 20 Sbjct: 242 DKLVERFKS--EEEFYNDRSFGLLYMLENHEQVKDKKNWTMALPLIGDVPKWSGVIEEYE 299

Query: 319 MSKNDEPAQQAELMAKTFLPNVNNLAYFSNEECKGWSKFDSESLFVGDDERNARCVIGID 378
 +++ DPA Q + +A LP+ + YF+ ++ K +F+ S+F R +GID
 25 Sbjct: 300 LAQGDPAQKFLAFNMGLPMQDTAYYFTPQDTK--LTFENLSVF----NKNRTYVIGID 352

Query: 379 LSDVNDICISIFMVVRGEERHYLNKKFMPRHTIETLPKELRDKYTEWELSGMLHVHLDY 438
 LS + D+ ++SF+ + + F R E L E ++ +TE+ G L + + + Y
 30 Sbjct: 353 LSLIGDLTAVSFVCELEGKTYSHTLTFSVRSQYEQLDTEQQELWTEFVDRGELILLDTEY 412

Query: 439 NDQAYIFEELRQFMSDNRLPVAVGVDYDRYNARELIRLFNDYYGDICHIPQTVK---SLS 495
 + + + F S +GYD L L Y+ D D + ++ S++
 Sbjct: 413 INVNDLIPYINDFRSKTGCRLRKIGYDPAVEILKGLIERYFFDKDGDNRRAIRQGFMSMN 472

Query: 496 NPLKVYKEKAKMGKIIFFDDPVATWNHANVRVKIDANNNIFPNKEKAKEKID 546
 + +K+ K K K+I + V W N VKI + + K+ K+KID
 35 Sbjct: 473 DYIKLLKSKLVENKLIHNQKVMQWALNNTAVKIGQSGDYMYTKKLEKDKID 523

No corresponding DNA sequence was identified in *S.pyogenes*.

40 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 813

A DNA sequence (GBSx0862) was identified in *S.agalactiae* <SEQ ID 2469> which encodes the amino acid sequence <SEQ ID 2470>. Analysis of this protein sequence reveals the following:

45 Possible site: 53
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 50 bacterial cytoplasm --- Certainty=0.3319(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

55 >GP:AAB41469 GB:L35061 orfL4 [Bacteriophage phi-41]
 Identities = 86/374 (22%), Positives = 166/374 (43%), Gaps = 38/374 (10%)

Query: 12 FARIFRPNNRKSTRTYLQRSISYWRNRSIYLDNIYNKISTDTAQLRFKHVKITRNPGGVD 71
 F+R N+ + + ++ Y S ++ NI+NKI+ + ++ F HVK ++ G D
 Sbjct: 10 FSRGKLNDTQRTAWQNEAVEY---TSAFVTNIHNKIANEITKVEFNHVKYKSDVGS 66

60 Query: 72 SMVWYEHSDLAEVLTVSPNPLEVPVVFWSNVTRAMLRDGVAVVVPWR--KNGRLVEIWL 129

-900-

```

      +++   SDL EVL S      + FW V + +L      + P + K G LV++ A
Sbjct: 67 TLISMAGSDLDEVLNWSSKGERNSMFEWQKVIKLLTTRYIDLYPIFDRKTGDLVDLLFA 126

Query: 130 KKTVTWTAESEVELMLDDVAVELEPLTDVWVFENPKLVNTAQLNQITELIDINLNALETEKLS 189
      + E + ++ +      N+ T ++D L + KL
Sbjct: 127 DNKKEYKPEELVRLISPFI-----NEDTSILDNALAGIQTKLE 165

Query: 190 DGNSSIRGFLKLPT---KAADHLKQQARDRVDSMLDLAKNGGIAYLEQGEEFQELSKDY 246
      G ++G LK+      D+ K +A + +M +++ G+ + E EL KDY
Sbjct: 166 QGK--MKGLLKINAFIDTDNDQEFKDKAMLTIKMQEMSNYGLTPTDNKTEIVELKKDY 223

Query: 247 STASKEELEFLKSQLYNAHGINEKLTCDYTEEQYRAYYSSVMKLYQRYVYSEEINRKYFT 306
      S +K+E++ +KS+L + +NE + ++EQ +Y+S + +E+ K +
Sbjct: 224 SVLNKDEIDLIKSELLTGYFMNENILLGTASQEQIYFYNSTIIPLLIQLEKELTYKLIS 283

Query: 307 KTAR--TQGN---KLLVFFDMADMISFKDLVEGGFKSKYAGLMNSNEFRETYLGLPGYE 360
      R +GN +++V + + K+L++ ++ + N+ +G E
Sbjct: 284 TNRRRVVKGMLYYERIIVDNQLFKFATLKLIDLYHENINGPIFTQNQLL-VKMGEQPIE 342

Query: 361 GGEVFETNLNAVRI 374
      GG+V+ NLNAV +
Sbjct: 343 GGEVYIANLNAVAV 356

```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 814

A DNA sequence (GBSx0863) was identified in *S.agalactiae* <SEQ ID 2471> which encodes the amino acid sequence <SEQ ID 2472>. This protein is predicted to be a prohead protease. Analysis of this protein sequence reveals the following:

```

Possible site: 25
>>> Seems to have no N-terminal signal sequence

```

```

----- Final Results -----

```

```

bacterial cytoplasm --- Certainty=0.3496(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAF31089 GB:AF069529 protease [Bacteriophage HK97]
Identities = 52/142 (36%), Positives = 73/142 (50%), Gaps = 11/142 (7%)

Query: 21 FEAVASTYDNTDREGDVMAKGCFDNTLKSKA-VVPMCLNHDR-NCVIGKHE-LSVDEKGL 77
      FE YAS ++NTD +GD++ G F N L ++ V M NH +GK + L+ DEKGL
Sbjct: 26 FEGYASVFNNNTSDSGDIILPGAFKNALANQTRKVAMFFNHKTWELPVGKWDSLAEDEKGL 85

Query: 78 RTRSTFNLSDEPAKKTIDLMKMGALDSLSIGFFI--KDYEPIDAKQPYGGWIFKEVE-IF 134
      R A M+ G ++ +S+GF + DY I G IFK ++ +
Sbjct: 86 YVRGQLTPGHSGAADLKAAMQHGTVEGMSVGFSVAKDDYTIIPT-----GRIFKNIQALR 140

Query: 135 EISVVTVPANPQATVDNIKEFD 156
      EISV T PAN QA + +K D
Sbjct: 141 EISVCTFPANEQAGIAAMKSVD 162

```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 815

A DNA sequence (GBSx0864) was identified in *S.agalactiae* <SEQ ID 2473> which encodes the amino acid sequence <SEQ ID 2474>. Analysis of this protein sequence reveals the following:

```

Possible site: 47
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2247(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 10155> which encodes amino acid sequence <SEQ ID 10156> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAC27185 GB:AF009630 16 [bacteriophage bIL170]
Identities = 70/249 (28%), Positives = 121/249 (48%), Gaps = 23/249 (9%)

Query: 51 LEQLKTDAESLVSQATA--IKETIAGLSDIEETEEELSK-AAKIIK-----EKQK 98
      L +LK + SL SQ +K I L ++E E+ Ls+ + +IIK EK K
Sbjct: 13 LAELKENNVSLKSGINGFEVKNAIEDLPK-VQELEKTLSENSIEIIKIENELNAQEEKPK 71

Query: 99 GNTPM-DYLKTKAAALDFVRILMDNEGSANSARKAWBANLVEKGV--TNLTkilPEPVL 155
      G M +++++ A +F +L N G + + AW A L E GV T+ T LP ++
Sbjct: 72 GKAKMTNFIESQNAVTEFFDVLKKNKSGKSE-IKNAWNAKLAENGVTITDTTfQLPRKLVE 130

Query: 156 AIQDAFTNYNGILN--HVS KDPRYAVRVALQTQVSAKGHKAGKTKKDEDFtFLDFTINS 213
      +I A N N + HV+ V + + ++A+ HK G+TK ++ T T+
Sbjct: 131 SINTALLNINPVFKVHVTVNGALLVSRsFDSS-ABEQVHKDGTQTEQAATLTIDTLEP 189

Query: 214 ATVY-IKYAFEYSDLKDDTTGAYFNYVMKELAQGFI-RTIERAVVIGDGKSN-SAEDKIT 270
      VY ++ E + + +N ++ EL Q + + ++ A+V GDG + + DK
Sbjct: 190 VMVYKLQSLAERVKRLQMSYSELYNLIVAEltQAIVNKIVDLALVEGDGSGNGFKSIDKEA 249

Query: 271 EIKSIAEET 279
      ++K I + T
Sbjct: 250 DVKKIKKIT 258

```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 816

A DNA sequence (GBSx0865) was identified in *S.agalactiae* <SEQ ID 2475> which encodes the amino acid sequence <SEQ ID 2476>. Analysis of this protein sequence reveals the following:

```

Possible site: 39
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3068(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 817

A DNA sequence (GBSx0866) was identified in *S.agalactiae* <SEQ ID 2477> which encodes the amino acid sequence <SEQ ID 2478>. Analysis of this protein sequence reveals the following:

Possible site: 56
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.0437(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 818

A DNA sequence (GBSx0867) was identified in *S.agalactiae* <SEQ ID 2479> which encodes the amino acid sequence <SEQ ID 2480>. Analysis of this protein sequence reveals the following:

Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3181(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10153> which encodes amino acid sequence <SEQ ID 10154> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 819

A DNA sequence (GBSx0869) was identified in *S.agalactiae* <SEQ ID 2481> which encodes the amino acid sequence <SEQ ID 2482>. This protein is predicted to be a major structural protein. Analysis of this protein sequence reveals the following:

Possible site: 29
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3364(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-903-

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAA74331 GB:L33769 unidentified ORF28; putative [Bacteriophage
      bIL67]
Identities = 55/201 (27%), Positives = 84/201 (41%), Gaps = 18/201 (8%)
5
Query:  9  EVTHGNANGF-YAKIAKTDAGALDLQKPYPFPTGLRSTSFETSQESNAYYAD-NVEHVRLQ 66
      E+THG  G  +  +  +  G          P  GLR  ++  QE+  +YA  N  +  +
Sbjct:  8  ELTHGLGYGVVFTDLTGSKTGI-----PIAGLRGIETDSKQENKNFYAGFNAPYRTIA 60

10
Query: 67  GKKSTEGSITTYQIPKQFMIDHLGKKLTNSTPPALIDTGVNIN-FIWGYAETVTDEFGAE 125
      G K T+  + +Y +P  F   LG   S   L D  N  + + YAE  D+ G
Sbjct: 61  GAKDTQIKVKSIDLDDFATHALG---FGSVQGFLTDDVANYKPYGFAYAEERYRDDGTG 117

15
Query: 126 IEEFHIWTVKASAPKSGTSTDETSATPKIEIIPCTASPNNFIVDSEKKPVSEIVWRDDS 185
      +   + +V+A+ P  +   DE S T KE E   T +  +F +  +K+   +   D
Sbjct: 118 YKA-TFYPVQATTPSDTAEADEESPTGKEYEHEATVTTGDFTLGDKKRLFVKFKVSDTE 176

Query: 186 KGT-VRGK---FDKLEADKSP 202
      T   GK   F  KLF  D   P
20
Sbjct: 177 LATGTSGKALAFKKLFTDLKP 197

```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

25 Example 820

A DNA sequence (GBSx0870) was identified in *S.agalactiae* <SEQ ID 2483> which encodes the amino acid sequence <SEQ ID 2484>. Analysis of this protein sequence reveals the following:

```

Possible site: 61
>>> Seems to have no N-terminal signal sequence
30
----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2531(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
35

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

40 Example 821

A DNA sequence (GBSx0871) was identified in *S.agalactiae* <SEQ ID 2485> which encodes the amino acid sequence <SEQ ID 2486>. Analysis of this protein sequence reveals the following:

```

Possible site: 22
>>> Seems to have no N-terminal signal sequence
45
----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2972(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
50

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 822

A DNA sequence (GBSx0872) was identified in *S.agalactiae* <SEQ ID 2487> which encodes the amino acid sequence <SEQ ID 2488>. Analysis of this protein sequence reveals the following:

Possible site: 49
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3860(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 823

A DNA sequence (GBSx0873) was identified in *S.agalactiae* <SEQ ID 2489> which encodes the amino acid sequence <SEQ ID 2490>. Analysis of this protein sequence reveals the following:

Possible site: 16
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -14.22 Transmembrane 605 - 621 (569 - 631)
INTEGRAL Likelihood = -8.12 Transmembrane 583 - 599 (569 - 604)

----- Final Results -----
bacterial membrane --- Certainty=0.6689(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAB70053 GB:AF011378 unknown [Bacteriophage sk1]
Identities = 159/709 (22%), Positives = 285/709 (39%), Gaps = 112/709 (15%)

Query: 128 SILNLNKELDNVAKELDIVNQKLELDPDNVELAEQKMKLLGKQSELAGDKVQELKKKQAA 187
S+ +N + + E + L+LDP N + Q K L Q L+ DK +LK++ ++
Sbjct: 21 SLKGVNTAMSGLRGEAKNLRDALKLDPTNTDKMAQLQKNLQTQLGLSRDKATKLKQELSS 80

Query: 188 LGDEK-IGTEEWRLQNEIGQAEVEVLKIDRAMDILGESSRSATGDI--KEATSYLRADV 244
+ G ++W QL ++G AE + +++ + + + S + DI K T + + +
Sbjct: 81 VDKSSPAGQKKWLQLTRDLGTAETQANRLEGEIKQVEGAISSGSWDIDAKMDTKGVNSGI 140

Query: 245 MMDVADKAG-----QIGQKMDVAGKMTVDWSEIDEALDVTITKTGLTGD----- 289
+ +G QIG V A + W + +A+DT L
Sbjct: 141 DGMKSRFSGRLREIAVGVFVRQIGSSAVSAVGNGLKGW--VSDAMDTQKAMISLQNTLKFPG 198

Query: 290 -----ALAELEQEIADKDIATG-----MPTSFQAGD----AVGEL-----NTQFGLT 326
+Q +AKD + T+F GD AVG+ N FG T
Sbjct: 199 NGQDFDYVSKSMQTLAKDTNANTEDTLKLSTTFIGLGD SAKTAVGKTEALVKANQAFGGT 258

Query: 327 GEKLKSASELL-----IKYAEINE-TD-----ISSAISAKQAEIAYG--LTAE 367
GE+LK + + IN+ TD + S+ + A++ YG +A
Sbjct: 259 GEQLKGVVQAYGQMSASGKVSANINQLTDNNTALGSALKSTVMEMNPALKQYGSFASAS 318

Query: 368 DLGMV----LDNVTKAAQDTGQSVDITVQKAIDGAPQIKGLGLSFEEGA-----ALIGK 417

-905-

+ G + LD + G T + A D + L L A ++I K
 Sbjct: 319 EKGAI SVEMLDKAMQKLGAGGGAVTTIGDAWDSFNETLSLALLPTLDALTPIISSIIDK 378
 Query: 418 FEKSGVDSSAALSSLSKAAVIYAKD--GKTLTDGLNETVSAIQNSTSET--EALSIASEI 473
 5 G + AL S+ K Y K+ G +G ++S I + T LSI ++
 Sbjct: 379 MAGWGESAGKALDSIVK----YVKELWGALEKNGALSSLSKIWDGLKSTFGSVLSIIGQL 434
 Query: 474 FGSKAAPRMVDIAIQGAFSFDDLAEEAKSSSGTVSTTFDETLDPIDKLTQYSNQAKEGMA 533
 S A +D+ + A + ++ S T++ D I K+ ++ + E
 10 Sbjct: 435 IESFAG---IDS-----KTGESAGSVENVSKTIANLAKGLADVICKIADFAKKFSESKEG 485
 Query: 534 ELGGKLLLETVIPALEPLMGMLLESSVNWFTSLNETDQ-QTIVILGLVTTAVMMLLGAIAPI 592
 + L+T + AL + T+++ + QT + G + AI P
 15 Sbjct: 486 AID--TLKTSILVALTAGFVAFKIGSGIITAISAFKKLQTAIQAGTGMGAFNAVMAINPF 543
 Query: 593 VIAIGAIGAPVGI VVAIV-GAIAVITLIIQAIMNWGAITEWLQSTWDSCAA-----W 644
 V +GI +AAIV G + T W + ++L+S WD + W
 Sbjct: 544 VA-----LGIAIAIVAGLVYFFTQTETGKKAWASFVDFLKSAWDGIVSFFSGIGQW 595
 20 Query: 645 LSELWNTIVTATTAWSNFTAWLSGLWSSVSTGQSLWSSFTSSLSNIFSSLITGAQSLW 704
 +++W V A W W SG+ V Q++W+ T+ + +++++TG Q+ W
 Sbjct: 596 FADIWNGAVDGAKGIWQGLVDWFGSIVQGV---QNIWNGITTFFTTLWTTVVVTGIQTAW 651
 Query: 705 SSFTSTLSNLWGLVSTGSNLFNNLSSTISGIFNGILSTASNIWNSIKS 753
 + T + LW G+V+ + +F +SS ++G +N ++T + + KS
 25 Sbjct: 652 AGVTGFFTGLWDGIVNVVTTVFTTISSLVTGAYNWFVTFQPLISFYKS 700

There is also homology to SEQ ID 2492.

A related GBS gene <SEQ ID 8663> and protein <SEQ ID 8664> were also identified. Analysis of this
 30 protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 10
 McG: Discrim Score: -13.98
 GvH: Signal Score (-7.5): -2.78
 Possible site: 16
 35 >>> Seems to have no N-terminal signal sequence
 ALOM program count: 2 value: -14.22 threshold: 0.0
 INTEGRAL Likelihood = -14.22 Transmembrane 605 - 621 (569 - 631)
 INTEGRAL Likelihood = -8.12 Transmembrane 583 - 599 (569 - 604)
 PERIPHERAL Likelihood = 4.45 539
 40 modified ALOM score: 3.34
 *** Reasoning Step: 3
 ----- Final Results -----
 45 bacterial membrane --- Certainty=0.6689(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear)

The protein has homology with the following sequences in the databases:

50 27.1/51.7% over 981aa
 Bacteriophage sk1
 GP|2392838| unknown Insert characterized
 ORF00471(328 - 2976 of 3333)
 55 GP|2392838|gb|AAB70053.1||AF011378(9 - 990 of 999) unknown {Bacteriophage sk1}
 %Match = 7.3
 %Identity = 27.1 %Similarity = 51.7
 Matches = 164 Mismatches = 275 Conservative Sub.s = 149
 60 243 273 303 333 363 393 423 453
 MSINQBEKKTLNADLLSVMSD*KERRKSMTEFTEGLYVKFGANTVEFDRSVKGINTALSSSLKKDFNNINRQLKMDPDNV
 : : |: || :|: |:|:|:|:| |: : |: ||:| |
 MASNATFEVEIYGNITTKFENSLKGVNTAMSGLRGEAKNLRDALKLDPNTT
 10 20 30 40 50

-907-

5
10
15
20
25
30
35
40
45

```

                930      940      950
2886      2916      2946      2976      3006      3036      3066      3096
LGAIGQSIANTIMNTSNNINNVNFGVTTIREEADLNRLANVVGNRIABEELQRKTNLRGGMA*QKSMNLPLTV*KHLLSVMY
| : |  : :  : | : |  |  |  | : : | | :
LSSSGYGLSTNSVSSDNRTYNTFNVQGGAGQDVSNLARAIRREFELGRA
          960      970      980      990

```

SEQ ID 8664 (GBS58) was expressed in and purified from *E.coli* as a GST fusion. The purified protein is shown in lane 10 of Figure 193.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 824

A DNA sequence (GBSx0874) was identified in *S.agalactiae* <SEQ ID 2493> which encodes the amino acid sequence <SEQ ID 2494>. Analysis of this protein sequence reveals the following:

```

Possible site: 43
>>> Seems to have no N-terminal signal sequence

```

20
25
30
35
40
45

```

----- Final Results -----
          bacterial cytoplasm --- Certainty=0.2732(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 825

A DNA sequence (GBSx0875) was identified in *S.agalactiae* <SEQ ID 2495> which encodes the amino acid sequence <SEQ ID 2496>. Analysis of this protein sequence reveals the following:

```

Possible site: 18
>>> Seems to have no N-terminal signal sequence

```

35
40
45

```

----- Final Results -----
          bacterial cytoplasm --- Certainty=0.2467(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 10151> which encodes amino acid sequence <SEQ ID 10152> was also identified. A further related GBS nucleic acid sequence <SEQ ID 10935> which encodes amino acid sequence <SEQ ID 10936> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2497> which encodes the amino acid sequence <SEQ ID 2498>. Analysis of this protein sequence reveals the following:

45
50
55
60
65
70
75
80
85
90
95
100
105
110
115
120
125
130
135
140
145
150
155
160
165
170
175
180
185
190
195
200
205
210
215
220
225
230
235
240
245
250
255
260
265
270
275
280
285
290
295
300
305
310
315
320
325
330
335
340
345
350
355
360
365
370
375
380
385
390
395
400
405
410
415
420
425
430
435
440
445
450
455
460
465
470
475
480
485
490
495
500
505
510
515
520
525
530
535
540
545
550
555
560
565
570
575
580
585
590
595
600
605
610
615
620
625
630
635
640
645
650
655
660
665
670
675
680
685
690
695
700
705
710
715
720
725
730
735
740
745
750
755
760
765
770
775
780
785
790
795
800
805
810
815
820
825
830
835
840
845
850
855
860
865
870
875
880
885
890
895
900
905
910
915
920
925
930
935
940
945
950
955
960
965
970
975
980
985
990
995
1000

```

Possible site: 40
>>> Seems to have no N-terminal signal sequence

```

```

----- Final Results -----

```

-908-

bacterial cytoplasm --- Certainty=0.2136(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

5 An alignment of the GAS and GBS proteins is shown below.

Identities = 55/240 (22%), Positives = 92/240 (37%), Gaps = 20/240 (8%)

Query: 4 INELTIDGVKTSSFCDVLVETRPNVIVSSS--KTALLEHDSIGAVVQSNRHRGLIEKP 61
 I ++ ID TSS VL I+S S + +G S + N + I

10 Sbjct: 2 IPKVIIDDFDTSSIPNCVLTGYDVGDLSPSPFVENEAYGMNGTSRELESYNESKPTIM-- 59

Query: 62 YHITLIEPSDEEIYRFSALLNREKFW-LENEQEPTIRLWCYKVDSFEIGKDEFGAWVVDV 120
 +H++ + + I L + +FW + N ++ Y S +I +W V +

15 Sbjct: 60 WHLSTFDDAVNLINHLGLSKKIEFWHIPNS-----IYYDCLSVKINAVTMSSWRVTL 113

Query: 121 TFICHPTKFFKTTDIQTLTGNGVLRVQGSALAFPKITVVGQSASETSFTIGNQVIKLEKL 180
 +P ++ K + GNG + G+ + PKI V G + + TIG QV++L L

Sbjct: 114 KLALYPFRYAKGVSDVVIAGNGNINNAGNVFSEPKIVVEG--TGKGTLTIGKQVMEL-NL 170

20 Query: 181 SESLVMTNDPNDPSFKTASGKL---IKWAGDFITVDTAKGQNVGVVLGAGITSLKFETVW 237
 S + A G + I+ G F + G+ + GIT W

Sbjct: 171 SGKATIECKHGQQCVYDAEGNVKNSIRIRGSFFEIQPG---TQGIAVSGGITRTIISPRW 227

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 25 vaccines or diagnostics.

Example 826

A DNA sequence (GBSx0876) was identified in *S.galactiae* <SEQ ID 2499> which encodes the amino acid sequence <SEQ ID 2500>. This protein is predicted to be PblB. Analysis of this protein sequence reveals the following:

30 Possible site: 27
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.00 Transmembrane 952 - 968 (952 - 968)

35 ----- Final Results -----
 bacterial membrane --- Certainty=0.1001(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

40 >GP:AAG18640 GB:AY007505 PblB [Streptococcus mitis]
 Identities = 145/542 (26%), Positives = 255/542 (46%), Gaps = 52/542 (9%)

45 Query: 1 MLFLLDANVRTVKWNGIPLHEASSAIVKEETNGDFYLTVRYPITDSGIYQLIKEDMLIKS 60
 M++L + N PL+ A + + +E N + LT R+P +D +++ +KE+ +K+

Sbjct: 1 MIYLTNGNT-----PLNAYADKISQEANSTYQLTFRFPTSDV-LWEKLKEETFLKA 51

Query: 61 PVPVLGAQLFRIKKPIENDDSMDITAYHVSDDIMKRSITPVSVVGGQCAMALSQMVNQNAK 120
 + G Q F I + + + + A V + I P+S+ + ALS+ +

50 Sbjct: 52 D-DLHGEQDFVIFEVQKKHGYIQVYANQVMILLNNYVINPISLDRATGSTALSRFAGSI- 109

Query: 121 TGLGDFSFTSDIMDSRTFNTTETETLYSVLMDGKHSIVGTWEGELVRDNFALSIRKSRGA 180
 T FSF SDI + TFNT + + D KHSI+G W G+LVR + + + ++ G+

55 Sbjct: 110 TRYNTFSFFSDIDERHTFNTDSVNAMVAFTKD-KHSILGQWGGDLVRHGYQVRLKNGGS 168

Query: 181 DRGVVITTHKNLKSQRTKNSQGVVTRIARSTFKPDGAE-DEVTLRVSVDSPLINSYPY 239
 + + KNL SYQ +++ + TRI ++T K +G + + V VDSPL+N Y

Sbjct: 169 ENESLFMYKKNLSSYQHKSTSKSLKTRITFKATVKGEKAPDRKFSVVVDSPLVNKYSQ 228

60 Query: 240 INEKEYENNAETVED--LRKWAEAKFTNEGIDKVSDAIEIEAYELDGQVNLGDTVNLK 297

-909-

I E E N+ + ++ LRK+ E F D + D++EI+ V + D V+L
 Sbjct: 229 IYEDVIEVNDQDVKDEVGLRKYGEQYFRITLDCMLEDSLEIQVEGKSDVPVQIFDIVSLF 288
 Query: 298 SRKHSADLYKKAIAYEFNALTEEYISITFDDKPGVGGSGVSSGLSN-VADAILVASATAQ 356
 + D+ KK Y ++ + ++ +SI F G SG+S+ LSN V+DA+ + Q
 Sbjct: 289 HDRFKMDVRKKITKYTYSPMAKKLLSIGF----GQFKSGLSNMLSNAVSDAVKNETQHLQ 344
 Query: 357 D---VAVQRAVKNANAAFDFAEFGKTKTKINDDIEIAKAKVESFKSELSNRMDNQLLP--- 410
 + + +KNA+ AFD + + + D + AKAK E K L+ +D +
 Sbjct: 345 GQFATQLGKEIKNADLAFDRKKEELVNQFTDGLNAAKAKAEVKKSLTETIDQRFDRFDS 404
 Query: 411 -----LATEAKNLASQAQADLTRKEIELRAELNRQVTSTEAVK 448
 LA EAK ++ QA+ + K E + ++ + TS +
 Sbjct: 405 TGLNEIKQKABEALQVRGANITLLAQEAQISEQARQQMDSKFAEYKQSVDRFTSLSSQL 464
 Query: 449 ISLTNLSHNMDIIKQKALNDLRDAETRLKEADSVQQLATKRVEDKLTGLSTKLESFVGG 508
 NL +D + + ++L + E+D +++A + ++L + S +VGG
 Sbjct: 465 AGKANL---IDFQRVQEKSNLYERIIGSSESDIAEKVARMTLTNLQFQVEVGKYS-AVGG 520
 Query: 509 YN 510
 N
 Sbjct: 521 PN 522
 Identities = 47/183 (25%), Positives = 83/183 (44%), Gaps = 22/183 (12%)
 Query: 867 VTTLRVTKGTIPADWSPSPDDLKAYSDTKLEQTANEIKASVTSLDHKTLLKQTDITMTSEG 926
 +T L +GT W P+P+D +D LE T QT +T+
 Sbjct: 667 MTELDIFYEGTTDRRWQPAPEATLETKLEAT-----QTKLTLLQGS 709
 Query: 927 IVLRAGKTSNDVARAIGSYFKVTPDAIALFSSLIKVSNGMLVDGVSVSRKLVGAVETGH 986
 ++ TS A +I S T + I + + I++ G L+D +T+ + G
 Sbjct: 710 FAIQ-NLTS---AGSIVSQINATNNQILIEAEKIRLKGKTLTD-ELTAIDGYFKRLFVGE 764
 Query: 987 VKAGAITGVLLAAEAHTAEKLVQDAFFNKLMANDAYLKQLFAKSAFITQVQSVTISASQ 1046
 + ++ ++ +TA+KL +DQA +++D + L AK AFI +++SV +SA+
 Sbjct: 765 GTFAKLNAEIIIGSKTITADKLIMDQAMARLFVSSDIFTDTLAAKEAFINKLRSVVVSATL 824
 Query: 1047 ISG 1049
 G
 Sbjct: 825 FEG 827

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2501> which encodes the amino acid sequence <SEQ ID 2502>. Analysis of this protein sequence reveals the following:

Possible site: 25
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2445(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 143/552 (25%), Positives = 251/552 (44%), Gaps = 43/552 (7%)
 Query: 11 TVKWNGIPLHEASSAIVKEETNGDFYLTVRYPIITDSGIYQLIKEDMLIKSPVPVLGAQLF 70
 ++K+ PL A + +E N D+ L +YP LIK+ +++++ + G+QLF
 Sbjct: 3 SIKDDNTPLVAAFEDEITQEANSDYKLNFKYPKHE-YRPLIKKGIILEAD-DLHGSQLF 60
 Query: 71 RIKKPIENDDSMIDITAYHVSDDIMKRSITPVSVVGGCAMLQSMVQNAKTGLGDFSFTS 130
 RI + + +++ A V+DD+ +I +SV +S++ + K FSF S
 Sbjct: 61 RIFEITKRHGYINVYANQVADDLNGYAITDISVDRVQGMVTMSELAGSIKRE-HPFSFSS 119
 Query: 131 DIMDSRTFNTTETETETLYSVLMDGKHSIVGTWEGELVRDNFALSIRSRGADRGVVITTHK 190
 DI TFN ++ + L +GKHSI+G W GELVR+ + +++ + G D + K
 Sbjct: 120 DIDGRHTFNQSDVSVM-DALANGKHSIMQWGGELVRNKYQINLLKKAGKDTETLFMYKK 178

-910-

Query: 191 NLKSYQRTKNSQGVVTRIH-----ARSTFKPDG-----AEDEVTLRVSVDSPLI 234
 NLKSY+ T +G+V+ +H + DG + + T+RVSV+S L
 Sbjct: 179 NLKSYEETDTIKGLVLSILHLVAEVEEHEVETREASDGNIGHSESPKKTIRVSVESKLLK 238

5 Query: 235 NSYPYINEK--EYENNAETVEDLRKWAEAKFTNEGIDKVSDAIEIAYELDGQVVNLGD 292
 +++P I EK + ++ + +T EDL + + F D ++++I+ V L D
 Sbjct: 239 DTHPIIVEKTIKVDQDVKTEEDLLAYGKKYFEKTLCDIPGNSLKIDVTNNYEGAVRLFED 298

10 Query: 293 TVNLKSRKHSADLYKKAIAIYEFNALTEEYISITFDDKPGVGGSGVSSGLSNVADAILVAS 352
 T + + DL + Y F + SI F G + ++ +SN D + S
 Sbjct: 299 TAIVFHELYDRDLRMQITGYRFAPMANRLKSTIF----GEIKTNLAKQISNQIDNKVAES 354

15 Query: 353 ATAQDVA---VQRAVKNANAFAEFGKTKTKINDDIEIAKAKVESFKSELSNR-MDNQ 407
 D A +Q+ + NAN FD + K + +I D I+ A+A E +E++ + ++ +
 Sbjct: 355 TAQHDAAFEAKLQKQIDNANRIFDTKEAKLREEIEDGIKKAENAEVVKVAEVNAKVLAE 414

20 Query: 408 LLPLATEAK----NLASQAQADLTRKEIELRAELNRQVTSTEAVKISLTNLSHMDIIK 462
 L A + + + A + D +K E R L + + +L + D +
 Sbjct: 415 ELAKAVDERLKKFLSDADTKQDFDKKLEEFRTSLKDLEVDEKQIDDALAKAGFSKDSLA 474

Query: 463 QKALNDLRDAETRLKEADSVQQL-ATKRVEDKLTGLSTKLESFSGVGGYNYVIDGGEPKEL 521
 +ET A+ V T ++L G + K+ +F GY + GE E
 Sbjct: 475 DIKAKLEDTSETATVTANIVGSTGGTFYNNRRLDGDTDKVITFE-QGYIDIAHNNEGGEF- 532

25 Query: 522 MANFYGKTYDIN 533
 GKTY I+
 Sbjct: 533 ----EGKFTYTIS 540

A related GBS gene <SEQ ID 8665> and protein <SEQ ID 8666> were also identified. Analysis of this
 30 protein sequence reveals the following:

Lipop Possible site: -1 Crend: 3
 SRCFLG: 0
 McG: Length of UR: 11
 Peak Value of UR: 1.54
 35 Net Charge of CR: 1
 McG: Discrim Score: -3.43
 GvH: Signal Score (-7.5): -5.44
 Possible site: 58
 >>> Seems to have no N-terminal signal sequence
 Amino Acid Composition: calculated from 1
 40 ALOM program count: 1 value: -0.00 threshold: 0.0
 INTEGRAL Likelihood = -0.00 Transmembrane 897 - 913 (897 - 913)
 PERIPHERAL Likelihood = 1.48 932
 modified ALOM score: 0.50
 45 icml HYPID: 7 CFP: 0.100
 *** Reasoning Step: 3
 ----- Final Results -----
 50 bacterial membrane --- Certainty=0.1001(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

55 32.8/53.9% over 503aa
 EGAD|33685| hypothetical protein Insert characterized
 EGAD|71773|76294 hypothetical protein { } Insert characterized
 SP|P15317|YHYA_BPH44 HYPOTHETICAL 65 KDA PROTEIN IN HYALURONIDASE REGION. Insert
 characterized
 60 GP|215054|gb|AAA98102.1| M19348 ORF {Streptococcus pyogenes phage H4489A} Insert
 characterized
 PIR|B30566|B30566 hypothetical protein - phage H4489A Insert characterized
 ORF00870 (1957 - 3777 of 4272)

EGAD|33685|35003(37 - 540 of 593) hypothetical protein {*Streptococcus pyogenes*}
EGAD|71773|76294 hypothetical protein { } SP|P15317|YHYA_BPH44 HYPOTHETICAL 65 KDA PROTEIN
IN HYALURONIDASE REGION. GP|215054|gb|AAA98102.1||M19348 ORF {*Streptococcus pyogenes* phage
H4489A} PIR|B30566|E30566 hypothetical protein - *Streptococcus pyogenes* phage H4489A

Matches = 137 Mismatches = 175 Conservative Sub.s = 88

1749 1779 1809 1839 1869 1899 1929 1959
 TRLKEADSVQQLATKRVEDKLTGLSTKLESFSGGYNVIDGGPEKELMANFYGKTYDINPQLLERTSQATLSFSYEABS
 :: : | : | :
 MSRDPTYTINEHDLISFADGRFYVTFKADKSSSETVRIN
 10 20 30

2184 2214 2244
-----SKVKLERGTVAIDWNNRDETLKASFAEYKQTVDE-----
 | : | : | : : | | : | | : |
SAEQILLQVKSIDDERYSKFQETLNGIKQTVKSSESVESARTQLASMFDSRISGLDGTKYSRLSQITDLSRLDDGVGNYS
 130 140 150 160 170 180 190

```

2271      2301      2331      2361      2388      2418      2448
-----NLANLRTSTETLAAQLTSAESSIRQTSESFSNRLVLSLEY-KDSEPNNRASKRYFEASKSETAK
          :: | : : | | : | | | : | : | : | : | : |
TLSQKVGSGIDLRVSNAAANDVSRLSQTAAGLQSQTNA---NQNYSSLSQTVOGLQTTVRDNQSNATSRI-----
          210          220          230          240          250          260

```

```

      2478           2838           2868           2898           2928           2958           3009
QLSALRTEVN-----SFVANNANFRANSLKIRFTDSQLKFRVTTLRVTKGTIPADWSPSPDDLK-AYSDT--KLEQTANEI
                :||   ::| :||||   :   : | |: |   :   ||
-----NQLSDLIST-KVTKGDVETTIAQSYDKIAFAIRDKLPAASKMTGSEI
              270          290          280          300

```

```

3039      3069      3099      3129      3159      3189      3213      3243
KASVTSLSLHKTLKQTDITMTSEGI VLRAGKTSNDVARAIGSYFKVTPDAIALFSSLIKVSG-NMLVDG-SVTSRKLVTGA
:                                     ||| | : : | | : : || | | : : |
IS-----A-INLDRSGVKITGKNITLDGNSYISNAVIKDA
                        320      330      340

```

[illegible]

3501 3537 3567 3624 3648 3678
AMEIQMSGQILYYTD-----QAALKRVLSGYPTQFVKPATGTVSG-KGNAGVTVIG--SNRYGTESINDGGFGVGVR
| :|| | : | || | | | | : | | :| : || | : | : | |
ASRWDLNSANIDFNRDATINFNSKNNAVRK-SGTNTAFVFHSNATPKGYRGSAIYGITSSGGDGIDSASSGRFCGVRR

440 450 460 470 480 490 500

-----AWNGSNIDSLDLVGDEIRLASSAFDNSDGWDVRTLDGLKITPHNRAAERNRSRIEVDWILKNGSYSSLRD
: : : | : | | : : | : | : : : :
FFRYAEGLQHTAKVQDAEIYGDII-VFSDDFNIDRGFKMRPSLMPKMVDLNKMYYQAILGRCLWHANNTAWSWNFDTRS

520 530 540 550 560 570 580

Score = 87.8 bits (214), Expect = 4e-19
Identities = 88/273 (32%), Positives = 133/273 (48%), Gaps = 47/273 (17%)

-912-

5
Query: 370 AINLSRGVQIAGKNIALDGNTT----VNGAF-----GAKLGEFI-----KLRAD 410
AI L S ++++G N+ +DG+ T V GA GA G + KL+ D
Sbjct: 897 AIALFSSLIKVSG-NMLVDGSVTSRKLVTGAVETGHVKAGAITGVLLAAEAVTAEKLKVD 955

10
Query: 411 QIIGGTIDANKINVINLKASSIVGLDANFIKARISYAIT-DLECKVIKARNGAMTIDLQ 469
Q + AN + L A S FI S I+ + G VIK N AM I +
Sbjct: 956 QAFFNKLMANDAYLKQLFAKSA-----FITQVQSVTISASQISGGVIKALNNAMETIQMN 1009

15
Query: 470 SGQINHYTNESAMRRIDSSSTASQFIKMTKSGFISEIGNMQAAMTVIGSNSDGSSENHENKT 529
SGQI +YT+++A+++R+ S +QF+K +G +S GN A +TVIGSN G+E+ +
Sbjct: 1010 SGQILYYTDQAALKRVLSGYPTQFVKFA-TGTVSGKCN--AGVTVIGSNRYGTESTNDGG 1066

20
Query: 530 FGGIRIWNKGSSYQSTS FVELVGN--RVAIYGNKNRSPWLFDDSTTSGYAYLIPQNDRGIK 587
F G+R WNG + ++LVG+ R+A N W + SG + P N
Sbjct: 1067 FVGVRWNG----SNIDSLDLVGEIRLASSAFDSDGWDVRTLDSGLK-ITPHN----- 1116

25
Query: 588 HVIGRADRKIDQIHVGDIYV-QGERVAMMLKDL 619
RA + +I VGD+++ +G L+D+
Sbjct: 1117 ----RAABRNSRIEVDVWILKNGSYSSLRDI 1145
Score = 31.3 bits (69), Expect = 0.038
Identities = 34/151 (22%), Positives = 62/151 (40%), Gaps = 13/151 (8%)

30
Query: 160 QNADKKLSASYQLGIDGLKATMRSDKIGLQAEIQTTAQGLYQRYDNEIRKLSAKITTTSS 219
Q A K +A++ K + D +A++++ L R DN++ L+ + +S
Sbjct: 306 QRAVNANAAFPDAEFGKTKTKINDDIEIAKAKVESFKSELSNRMDNQLLEPLATEAKNLAS 365

35
Query: 220 GTTEAYESKLDGLRAEFTH---SNQGMRVELES-----KISGLQSTQQATARQISQE 268
K LRAE S + +++ L + K L + A R + +
Sbjct: 366 QAQADLTRKEIELRAELNRQVTSTEAVKISLTNLSHNMDI IKQKALNDLRDAETR-LKEA 424

Query: 269 ISNREGAVSRVQQGLDSYQRRLQS-AEGNYN 298
S ++ A RV+ L +L+S + G YN
Sbjct: 425 DSVQQLATKRVEDKLTGLSTKLESFSGVGGYN 455

SEQ ID 8666 (GBS202) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 50 (lane 5; MW 132kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

40 Example 827

A DNA sequence (GBSx0877) was identified in *S.galactiae* <SEQ ID 2503> which encodes the amino acid sequence <SEQ ID 2504>. This protein is predicted to be nuclear/mitotic apparatus protein. Analysis of this protein sequence reveals the following:

45 Possible site: 22
>>> Seems to have no N-terminal signal sequence

50 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.2847(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

55 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 828

A DNA sequence (GBSx0879) was identified in *S.agalactiae* <SEQ ID 2505> which encodes the amino acid sequence <SEQ ID 2506>. Analysis of this protein sequence reveals the following:

Possible site: 23
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3420(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 829

A DNA sequence (GBSx0880) was identified in *S.agalactiae* <SEQ ID 2507> which encodes the amino acid sequence <SEQ ID 2508>. Analysis of this protein sequence reveals the following:

Possible site: 13
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -7.54 Transmembrane 10 - 26 (2 - 28)

----- Final Results -----
 bacterial membrane --- Certainty=0.4015(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB07984 GB:Z93946 hypothetical protein [bacteriophage Dp-1]
 Identities = 67/136 (49%), Positives = 91/136 (66%)

Query: 1 MPPWLIDSTVVVAMVTVLGGLFSTIITTSANRKDQLIKHQYEDIKEDLSGLIDKVKITIDH 60
 MP WL D+ V+ ++T G+ + ++ K K EDI LS L +V ID
 Sbjct: 1 MPMWLNDTAVLTITITACSGVLTLLNKLFEWKSNAKSVLEDISTTLSTLKQQVDGIDQ 60

Query: 61 TTTETKKISEITKDGTCLKIQRRLFHDLTKEISQGYTTIEHFRELSILFESYQLLGGNGE 120
 TT +++ +DGT KIQRYRL+HDL +E+ GYTT++HFRELSILFESY+ LGGNGE
 Sbjct: 61 TTVAINHQNDVIQDGTCKIQRRLYHDLKREVITGYTTLDHFRELSILFESYKNLGGNGE 120

Query: 121 IEALFEKFKQLPIED 136
 +EAL+EK+K+LPI E+
 Sbjct: 121 VEALYEKYKKLPIREE 136

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 2508 (GBS118) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 32 (lane 5; MW 42kDa).

GBS118-GST was purified as shown in Figure 198, lane 8.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 830

A DNA sequence (GBSx0882) was identified in *S.agalactiae* <SEQ ID 2509> which encodes the amino acid sequence <SEQ ID 2510>. Analysis of this protein sequence reveals the following:

```

5   Possible site: 53
   >>> Seems to have a cleavable N-term signal seq.

   ----- Final Results -----
   bacterial outside --- Certainty=0.3000(Affirmative) < succ>
10  bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
   bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

15 A related GBS gene <SEQ ID 8667> and protein <SEQ ID 8668> were also identified. Analysis of this protein sequence reveals the following:

```

   Lipop: Possible site: -1   Crend: 5
   McG: Discrim Score:      6.58
   GvH: Signal Score (-7.5): -0.49
   Possible site: 53
20  >>> Seems to have a cleavable N-term signal seq.
   ALOM program   count: 0 value: 12.15 threshold: 0.0
   PERIPHERAL Likelihood = 12.15      84
   modified ALOM score: -2.93

25  *** Reasoning Step: 3

   ----- Final Results -----
   bacterial outside --- Certainty=0.3000(Affirmative) < succ>
30  bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
   bacterial cytoplasm --- Certainty=0.0000(Not Clear)

```

SEQ ID 2510 (GBS56) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 17 (lane 8; MW 9.9kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 21 (lane 10; MW 34.9kDa).

35 GBS56-GST was purified as shown in Figure 195, lane 7.

Example 831

A DNA sequence (GBSx0883) was identified in *S.agalactiae* <SEQ ID 2511> which encodes the amino acid sequence <SEQ ID 2512>. Analysis of this protein sequence reveals the following:

```

40  Possible site: 40
   >>> Seems to have a cleavable N-term signal seq.

   ----- Final Results -----
   bacterial outside --- Certainty=0.3000(Affirmative) < succ>
   bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
45  bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for
50 vaccines or diagnostics.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 832

A DNA sequence (GBSx0884) was identified in *S.galactiae* <SEQ ID 2513> which encodes the amino acid sequence <SEQ ID 2514>. This protein is predicted to be N-acetylmuramoyl-L-alanine amidase. Analysis of this protein sequence reveals the following:

Possible site: 53
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.0342(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB07986 GB:Z93946 N-acetylmuramoyl-L-alanine amidase
[bacteriophage Dp-1]
Identities = 96/141 (68%), Positives = 118/141 (83%)

Query: 1 MEINTEIAIAWMSARQKVSYSMDYRDGPNSYDCSSSVYYALRSAGASSAGWAVNTEYMH 60
M ++ E +AWM AR+G+VSYSMD+RDGP+SYDCSSS+YYALRSAGASSAGWAVNTEYMH
Sbjct: 1 MGVDIEKGVAVWQARKGRVSYSMDFRDGPDSYDCSSSMYYALRSAGASSAGWAVNTEYMH 60

Query: 61 DWLIKNGYELIAENVWDNAVRGDIWGMGRGHSSGAGGHVVMFIDPENIIHCNWNANGIT 120
WLI+NGYELI+EN W+A RGDI IWG +G S+GAGGH MFID +NIIHCN+A +GI+
Sbjct: 61 AWLIENGVELISENAPWDAKRGDIFIWGRKGASAGAGGHTGMFIDSDNIIHCNYAYDGIS 120

Query: 121 VNNYNQTAASGWMYCYVYRL 141
VN++++ +G Y YVYRL
Sbjct: 121 VNDHDERWYYAGQPYYYVYRL 141

No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8669> and protein <SEQ ID 8670> were also identified. Analysis of this protein sequence reveals the following:

RGD motif 81-83

The protein has homology with the following sequences in the databases:

58.2/72.9% over 182aa

GP|1934766| N-acetylmuramoyl-L-alanine amidase {bacteriophage Dp-1} Insert characterized

ORF00875(301 - 1044 of 2004)
GP|1934766|emb|CAB07986.1||Z93946(1 - 183 of 296) N-acetylmuramoyl-L-alanine amidase
{bacteriophage Dp-1}
%Match = 15.5
%Identity = 58.2 %Similarity = 72.8
Matches = 107 Mismatches = 49 Conservative Sub.s = 27

234	264	294	324	354	384	414	444
LQKYNHMSDDDLTLFVESAVKQMDAWKE*	PMEINTEIAIAWMSARQKVSYSMDYRDXPNSYDCSSSVYYALRSAGAS						
	:: : ::: :: : :						
	MGVDIEKGVAVWQARKGRVSYSMDFRDGPDSYDCSSSMYYALRSAGAS						
	10	20	30	40			

474	504	534	564	594	624	654	684
SAGWAVNTEYMHDWLIKNGYELIAENVWDNAVRGDIWGMGRGHSSGAGGHVVMFIDPENIIHCNWNANGITVNNYNQTA							
SAGWAVNTEYMHAWLIENGVELISENAPWDAKRGDIFIWGRKGASAGAGGHTGMFIDSDNIIHCNYAYDGISVNDHDERW							

-916-

```

      60      70      80      90     100     110     120
714      744      774      804      834      864      894      924
AASGWMYCYVYRLKSGASTQGKSLDTLVKETLAGNYNGEARKAVLGNQYEAVMSVINGKTTTNQKTVDQLVQEVIAQKH
5  :| | |||| :
YYAGQPFYYVYRLTNA-----
      140

954      984     1014     1044     1074     1104     1134     1164
GNGEARKKSLGSQYDAVQKRVTLELLKKQSEPFKAQEVNKPETETKTSQTELTGQATATKEEGDLSFNGTILKKAVLDKIL
10 | : :| || || | : | : || | : : :| : : | :
-NAQPAEKKLGNQKDATGFWYARANGTYPKDEFEYIEENKSWFYFDDQGYMLAEKWLKHTDGNWYWFDRDGYMATSWKRI
      160      170      180      190      200      210      220

```

SEQ ID 8670 (GBS302) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell
 15 extract is shown in Figure 50 (lane 6; MW 55kDa).

The GBS302-His fusion product was purified (Figure 205, lane 6) and used to immunise mice. The
 resulting antiserum was used for FACS (Figure 302), which confirmed that the protein is immunoaccessible
 on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 20 vaccines or diagnostics.

Example 833

A DNA sequence (GBSx0885) was identified in *S.agalactiae* <SEQ ID 2515> which encodes the amino
 acid sequence <SEQ ID 2516>. Analysis of this protein sequence reveals the following:

```

Possible site: 38
25 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.1509(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
30      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for
 35 vaccines or diagnostics.

Example 834

A DNA sequence (GBSx0886) was identified in *S.agalactiae* <SEQ ID 2517> which encodes the amino
 acid sequence <SEQ ID 2518>. Analysis of this protein sequence reveals the following:

```

Possible site: 19
40 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.1264(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
45      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:CAB13473 GB:Z99112 similar to hypothetical proteins [Bacillus subtilis]
Identities = 25/68 (36%), Positives = 41/68 (59%)

```

-917-

Query: 4 IENLIIAIVKPLISQPDQLTIKIQDGPEFLEYHLDLDTQDIGRVIGKKGRITTAIRSIVY 63
 +E+LI+ IV PL+ PD + + ++ + + L + D G+VIGK+GRT AIR+ V+
 Sbjct: 6 LEDLIVHIVTPLVDHPDIRVIREETDQKIALRLSVHKSDTGKVIQKQGRITAKAIRTAVF 65

5 Query: 64 SVPTQGGK 71
 + Q K
 Sbjct: 66 AAGVQSSK 73

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2519> which encodes the amino acid
 10 sequence <SEQ ID 2520>. Analysis of this protein sequence reveals the following:

Possible site: 19
 >>> Seems to have no N-terminal signal sequence

15 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1012(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

20 Identities = 72/79 (91%), Positives = 75/79 (94%)
 Query: 1 MDTIENLIIAIVKPLISQPDQLTIKIQDGPEFLEYHLDLDTQDIGRVIGKKGRITTAIRS 60
 MDTIENLIIAIVKPLISQPD LTIKI+D P+FLEYHLDLD QDIGRVIGKKGRITTAIRS
 Sbjct: 1 MDTIENLIIAIVKPLISQPDNLTIKIEDTPDFLEYHLDLDAQDIGRVIGKKGRITTAIRS 60
 25 Query: 61 IVYSVPTQGGKVRLLIIDEK 79
 IVYSVFT GKKVRL+IDEK
 Sbjct: 61 IVYSVPTLGGKVRLLVIDEK 79

30 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 vaccines or diagnostics.

Example 835

A DNA sequence (GBSx0887) was identified in *S.agalactiae* <SEQ ID 2521> which encodes the amino
 acid sequence <SEQ ID 2522>. This protein is predicted to be ribosomal protein S16 (rpsP). Analysis of
 35 this protein sequence reveals the following:

Possible site: 45
 >>> Seems to have no N-terminal signal sequence

40 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3654(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

45 >GP:BAB06202 GB:AP001515 ribosomal protein S16 (BS17) [Bacillus halodurans]
 Identities = 62/90 (68%), Positives = 73/90 (80%)
 Query: 1 MAVKIRLTRMGSKKKPFYRINVADSRAPRDGRFIEITVGTYNPLVAENQVTIKEERVLEWL 60
 MAVKIRL RMGSKK PFYR+ VADSR+PRDGRFIE +GTYNPL +V +KE+R L+W+
 50 Sbjct: 1 MAVKIRLKRMGSKKAPFYRVVVADSRSPRDGRFIEIGTYNPLTQPAKVELKEDRALDWM 60
 Query: 61 SKGAQPSDTPVRNLLSKAGVMTKFHDQKFSK 90
 KGA+PSDTPVRNL SKAG+M K H+ K K
 Sbjct: 61 LKGAKPSDTPVRNLFSKAGLMEKLNHAKNEK 90
 55

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2523> which encodes the amino acid
 sequence <SEQ ID 2524>. Analysis of this protein sequence reveals the following:

-918-

Possible site: 45

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.3654(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

10 Identities = 86/90 (95%), Positives = 89/90 (98%)

Query: 1 MAVKIRLTRMGSKKKPFYRINVADSRAPRDGRFIETVGTYNPLVAENQVTIKEERVLEWL 60
 MAVKIRLTRMGSKKKPFYRINVADSRAPRDGRFIETVGTYNPLVAENQ+TIKE+RVLEWL
 Sbjct: 1 MAVKIRLTRMGSKKKPFYRINVADSRAPRDGRFIETVGTYNPLVAENQITIKEDRVLEWL 60

15 Query: 61 SKGAQPSDITVRNLLSKAGVMTKFHDQKFSK 90
 SKGAQPSDITVRN+LSKAGVM KFHDQKFSK
 Sbjct: 61 SKGAQPSDITVRNLLSKAGVMAKFHDQKFSK 90

20 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 vaccines or diagnostics.

Example 836

A DNA sequence (GBSx0888) was identified in *S.agalactiae* <SEQ ID 2525> which encodes the amino
 acid sequence <SEQ ID 2526>. Analysis of this protein sequence reveals the following:

25 Possible site: 35

>>> Seems to have no N-terminal signal sequence

 INTEGRAL Likelihood = -11.09 Transmembrane 22 - 38 (16 - 42)
 INTEGRAL Likelihood = -7.64 Transmembrane 382 - 398 (375 - 402)
 INTEGRAL Likelihood = -7.59 Transmembrane 291 - 307 (284 - 317)
 30 INTEGRAL Likelihood = -4.94 Transmembrane 340 - 356 (335 - 366)

----- Final Results -----

 bacterial membrane --- Certainty=0.5437(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 35 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC24912 GB:AF012285 YknZ [Bacillus subtilis]

40 Identities = 161/417 (38%), Positives = 241/417 (57%), Gaps = 25/417 (5%)

Query: 1 MENWKFALSSILGHKMFRAFLTMLGIIIGVASVVLIMALGKGMKDSVTNEITKSQKNLQIY 60
 +EN + ALSS+L HKMR+ LITMLGIIIGV SV++++A+G+G + + I+ +++Y
 Sbjct: 4 LENIRMALSSVLAKHKRSILITMLGIIIGVSVIVVVAVGQGGEQMLKQISGPGNTVELY 63

45 Query: 61 YKTKEDQ-KNEDNFGAQQGAFMQGSDTNRKEPIIQESWLKKIAKEVDGVSQYVYNQTNAP 119
 Y +++ + N A+ F + K K ++G+ + +
 Sbjct: 64 YMPSEELASNPNAEAESTFTENDI-----KGLKGIEGIQVAVSTSESMSK 109

50 Query: 120 VAYLEKKAKTVNITGINRTYLGIKKFKIKSGRQFQEEEDYNQFSRVILLEEKLAQRLEQYN 179
 Y E++ + GIN Y+ + KI+SGR F + D+ +RV ++ +K+A+ LF
 Sbjct: 110 ARYHEETDAT-VNGINDGYMNVNSLKIESGRTFTDNDFLAGNRVGIISQKMAKELFDKT 168

 Query: 180 EAALNKVVTVKNSYLVGVYSDPEAGSGLYGSNSDGNAILTNTQLASEFGAKEAENIYF 239
 + L +VV + + ++GV +GL + + N + S FG + N+
 55 Sbjct: 169 -SPLGEVWINGQFVEIIGVLKKV---TGLLSFDLSEMYVFPN-MMKSSFGTSDFSNVSL 223

 Query: 240 HLNDVQSQRNIGKEIGKRLTDISHAKDGYDNDNMTSIVKSINTQVGIMTGIGAIAAIS 299
 + GKE + + D +H + Y +M I I IMT +IG+IA IS
 Sbjct: 224 QVESADDIKSAGKEAAQLVND-NHGTEDSYQVMNMEETAGIGKVTAIMTTIIGSIAGIS 282

60 Query: 300 LLVGGIGVMNIMLVSVTERTREIGLRKALGATRKRKILAQFLIESMVLITLGLIGLLLAY 359

-919-

LLVGGIGVMNIMLVSVTERTREIG+RK+LGATR +IL QFLIES+VLT++GGL+G+ + Y
 Sbjct: 283 LLVGGIGVMNIMLVSVTERTREIGIRKSLGATRQILTQFLIESVVLTLLIGGLVIGIGY 342

Query: 360 GGTMLIANAQDKITPS-VSLNVAIGSLIFSFIGIIFGLLPANKASKLNPIDALRYE 415
 5 GG L++ PS +S V G ++FS IG+IFG+LPANKA+KL+PI+ALRYE
 Sbjct: 343 GGAALVSAIAG--WPSLISWQVVCVGGVLFMSLIGVIFGMLPANKAAKLDPIEALRYE 397

There is also homology to SEQ ID 1350.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for
 10 vaccines or diagnostics.

Example 837

A DNA sequence (GBSx0889) was identified in *S.agalactiae* <SEQ ID 2527> which encodes the amino acid sequence <SEQ ID 2528>. This protein is predicted to be ABC transporter (ATP-bindingprot). Analysis of this protein sequence reveals the following:

15 Possible site: 52
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 20 bacterial cytoplasm --- Certainty=0.4080(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

25 >GP:BAB06841: GB:AP001517 ABC transporter (ATP-binding protein)
 [Bacillus halodurans]
 Identities = 131/218 (60%), Positives = 169/218 (77%)

Query: 8 LIRLHQIVKSYQNGDQKLQVLKNIDLTVYEGEFLAIMGPSGSGKSTLMNIIGLLDSPTSG 67
 +I+L ++ KS++ G + +++L IDL + G+FLAIMGPSGSGKSTLMNIIG LD PTSG
 30 Sbjct: 1 MIKLERVTKSFRVGTENVIELSAIDLEIASGDFLAIMGPSGSGKSTLMNIIGCLDQPTSG 60

Query: 68 DYSLNGKRVRELSQTKLAQVRNKEIGFVFQQFFLLSKLTALQNVELPLIYAGVPPKKRN 127
 Y +GK + S+ ++A++RN+ IGFVFQQF LL +LTALQNVELP++YAG+ K+R
 35 Sbjct: 61 RYMFQDKDLTNYSEQEIAKIRNRHIGFVFQQFHLLPRLTALQNVELPMVYAGMKKKERTE 120

Query: 128 LAKQFLDKVELRERMNHLPTLSEGGQKQVVAIARALVNSPSIILADEPTGALDTKTGEQI 187
 A L++V L ERM +LP LSGGQKQVVAIAR++VN P+IILADEPTGALDTKT E I
 Sbjct: 121 RAAHALERVGLAERTYLPNSLSEGGQKQVVAIARSIVNEPNIIILADEPTGALDTKTSETI 180

40 Query: 188 MQFLTELNQEGKTIIMVTHEPEIADYATRKIVIRDGEI 225
 M+ L LN EG TI +VTHEPEIA+Y + + +RDG+I
 Sbjct: 181 MELLCSLNNEGTTIALVTHEPEIAEYVQQTQVFVRDQI 218

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2529> which encodes the amino acid
 45 sequence <SEQ ID 2530>. Analysis of this protein sequence reveals the following:

Possible site: 52
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 50 bacterial cytoplasm --- Certainty=0.1739(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

55 Identities = 182/232 (78%), Positives = 207/232 (88%)

Query: 5 RKELIRLHQIVKSYQNGDQKLQVLKNIDLTVYEGEFLAIMGPSGSGKSTLMNIIGLLDSP 64

-920-

+K+L++L IVKSYQNGDQ L+VLK I+LTVYEGEFLAIMGPSGSGKSTLMNIIGLLD P
 Sbjct: 5 KKQLMQLSNIVKSYQNGDQVLKVLKGINLTVYEGEFLAIMGPSGSGKSTLMNIIGLLDRP 64
 Query: 65 TSGDYSINGKRVEELSQTCLAQVRNKEIGFVFQQFFLLSKLTALQNVELPLIYAGVPPKK 124
 TSGDY+L+ ++E L+ +LA+VRN EIGFVFQQFFLL+KLTALQNVELPLIYAGV K
 Sbjct: 65 TSGDYTLHNTKIEILNDRBLAKVRNDEIGFVFQQFFLLAKLTALQNVELPLIYAGVNVSK 124
 Query: 125 RKNLAKQFLDKVELRERMNHLPTELSGGQKQORVAIARALVNSPSIILADEPTGALDTKTG 184
 R+ AKQFL+KV L R+ HLP+ELSGGQKQORVAIARALVN PSIILADEPTGALDTKTG
 Sbjct: 125 RREQAKQFLEKVGGLRRRIKHLPSSELGGQKQORVAIARALVNDPSIILADEPTGALDTKTG 184
 Query: 185 EQIMQFLTTELNQEGKTIIMVTHEPEIADYATRKIVIRDEITADTTDSIRID 236
 +QIM+ LTELN+EGKTIIMVTHEPEIAD+ATRKI+IRDG+IT DTT S+ ID
 Sbjct: 185 QQIMELLTELNKEGKTIIMVTHEPEIADFATRKIIIRDGDIITDTTASVVID 236

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 838

A DNA sequence (GBSx0890) was identified in *S.agalactiae* <SEQ ID 2531> which encodes the amino acid sequence <SEQ ID 2532>. This protein is predicted to be ATP-binding cassette transporter-like protein. Analysis of this protein sequence reveals the following:

Possible site: 37
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -8.97 Transmembrane 17 - 33 (13 - 39)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.4588(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9965> which encodes amino acid sequence <SEQ ID 9966> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC24909 GB:AF012285 YknX [Bacillus subtilis]
 Identities = 104/391 (26%), Positives = 182/391 (45%), Gaps = 21/391 (5%)
 Query: 13 KKGAIISGLSVALIVVIGGFLWVQSQPNKSAVKTNKYKVFNVREGSVSSSTLLTGKAKANQ 72
 KK I G++V + + +G ++ + P + + +V E +SS+ ++ G K +
 Sbjct: 2 KKVWIGIGIAVIVALFVGINIRSAAPTSAGSAGKEVQTGSVEENEISSTVMVPGTLKFSN 61
 Query: 73 EQVYVFDANKGNRAITVTVKVGDKITAGQQLVQYDTTTAQAAAYDTANRQLNKVARQINNLIK 132
 EQVV+++A+KG + VK GDK+ G LV Y T Q + + QL + ++ +
 Sbjct: 62 EQVVFYEADKGTLEDIKVKEGDKVKKGTALVTY--TNEQLSLEKEQNQLTSSESRLQIDQ 119
 Query: 133 TTGSLPAMESSDQSSSSSQGGTQSTSGATNRLQNYQSQANASYNQQLQDLNDAYADAQ 192
 L A++S ++ G+ + R + Q + +L Q
 Sbjct: 120 IQEKLKALDSKERELEKQVGKKEAEKQIESERTELQMOKKTAEI-----ELKQTELQRQ 173
 Query: 193 ARVNKAQKALNDTVITSDVSGTVVEVNSDIDPASKTSQV---LVHVATEGKLQVQGTMSSE 249
 + N+ ++D + S++ GTV+ VN + ASK S + ++H+ L V G +SE
 Sbjct: 174 SLANR----VSDLEVKSIEGTVISVQ--EAASKSDIQEPVIHIGNPKDLVVSGLKSE 227
 Query: 250 YDLANVKKDQAVKIKSKVYPDKKEWEGKISYISNYPEAEANNDNSNNGSSAVNYKYKVDIT 309
 YD VKK Q V + S V K W+G +S + P+ + + + AV Y +V I
 Sbjct: 228 YDTLKVKKGQKVTILTSVIOGKTWKGTVSAVGLVPD-QQESAAAQGTQAVQYPLQVKIK 286
 Query: 310 SPLDALKQGFTVSVFV-VNGDKHLIVPTSSVINKDNKHFVWVYNDNSNRKISKVEVKIGKA 368
 L K GF + + + K +P+ +V +D+++V+ D K +V+VKIG+
 Sbjct: 287 GNLPEGKPGFKFIMNIETDKRKANTLPSKAVKKEDDQYYVYTVKDG--KAKRVDVKIGEV 344

-921-

Query: 369 DAKTQEILSGLKAGQIVVTNPSKTFKDGQKI 399

EI GL V+ NPS DG ++

Sbjct: 345 TDDLTEIKEGLTQDDQVILNPSDQVTDGMEV 375

5

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2533> which encodes the amino acid sequence <SEQ ID 2534>. Analysis of this protein sequence reveals the following:

Possible site: 42

>>> Seems to have an uncleavable N-term signal seq

10

INTEGRAL Likelihood = -9.61 Transmembrane 15 - 31 (11 - 36)

----- Final Results -----

bacterial membrane --- Certainty=0.4843(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

15

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAC24909 GB:AF012285 YknX [Bacillus subtilis]

Identities = 103/380 (27%), Positives = 180/380 (47%), Gaps = 21/380 (5%)

20

Query: 16 ITASVITLVLIITGIVLWKQQRNTLTADIAKEPYSTVSVTEGSIASSTLLSGTVKALSEE 75

I + +V + GI +++ T + A + T SV E I+S+ ++ GT+K +E+

Sbjct: 6 IGIGIAVIVALFVGINIRSAAPT--SGSAGKEVQTGSVEENEISSTVMVPGTLKFSNEQ 63

25

Query: 76 YIYFDANKGNDAITVTVKVGDDQVTTGQQLVQYNTTTAQSAYDTAVRSLNKIGRQINHLKTY 135

Y++++A+KG + VK GD+V +G LV Y T Q + + + N++ + N L+

Sbjct: 64 YVFEADKGTLEDIKVKEGDKVKKGTALVTY--TNEQLSLE---KEQNQLTSESRLQID 118

30

Query: 136 GVPVSTETNRDEATGEETTTTVQPSAQ-QNANYKQQLQDLNDAYADAQAEVNKAQIA-- 192

+ + E E+ + Q ++ + Q+Q Q E+ + +A

Sbjct: 119 QIQEKLKALDSKERELEKQVGKKEAKQIESERTELQMQKKTAEIELKQTELQRLQSLANR 178

35

Query: 193 LNDTVVISSVSGTVVEVNND-IDPSSKNSQTLVHVATEGQLQVKGTLTEYDLANVKVQGS 251

++D V S + GTV+ VN + S + ++H+ L V G L+EYD VK GQ

Sbjct: 179 VSDLEVKSEIEGTVISVNEAASKSDIQEPVIHIGNPKDLVVSGLSEYDTLKVKKGQK 238

40

Query: 252 VKIKSKVYSNQEWGTGKISYVSNYPTESNAGSTTFAGSTGAGSSTGATYDYKIDIISPLNQ 311

V + S V + W G +S V P + + + G+ Y ++ I L +

Sbjct: 239 VTLTSDVIQKKTWKTSAVGLVPDQQES-----AAAQGTQAVQVPLQVKIKGNLPE 291

45

Query: 312 LKQGFVTSVEVVNEAKQA-LVPLTAVIKKDKKHVWVYDDATGKAKKVEVTLGNADAQQQ 370

K GF + + + ++A +P AV K+D ++YV+T D GKAK+V+V +G

Sbjct: 292 GKPGKFKIMNIETDKRKANTLPSKAVKKEDDQYVYTVKD--GKAKRVDVKIGEVTDDLIT 349

Query: 371 EIHKGAVAGDIVIANPDKNI 390

EI +G+ D VI NP +

Sbjct: 350 EIKEGLTQDDQVILNPSDQV 369

An alignment of the GAS and GBS proteins is shown below.

50

Identities = 234/421 (55%), Positives = 301/421 (70%), Gaps = 19/421 (4%)

Query: 3 MSKRQNLGISKKGAIISGLSVALIVVIGGF-LWVQSQPNKSA--VKTNKYKFNVREGSVS 59

MSKR + I+ K +I+ + I+++I G LW Q + +A K Y +V EGS++

Sbjct: 1 MSKRKGKIKITTKTLITASVITLVLIITGIVLWKQQRNTLTADIAKEPYSTVSVTEGSI 60

55

Query: 60 SSTLLTGKAKANQEYVYFDANKGNRATVTVKVGDKITAGQQLVQYDTTTAQAAAYDTANR 119

SSTLL+G KA E+Y+YFDANKGN ATVTVKVG++T GQQLVQY+TTTAQ+AYDTA R

Sbjct: 61 SSTLLSGTVKALSEEYIYFDANKGNDAITVTVKVGDDQVTTGQQLVQYNTTTAQSAYDTAVR 120

60

Query: 120 QLNKVARQINNLKTTGSLPAMESSDQSSSSSQGGTQSTSGATNRIQQNYQSQANASYNQ 179

LNK+ RQIN+LKT G +PA+ S++ + + G+ T +T + +Q NA+Y Q

Sbjct: 121 SLNKIGRQINHLKTYG-VPVAV-STETNRDEATGEETTTTVQPS-----AQQNANYKQ 170

Query: 180 QLQDLNDAYADAQAEVNKAQKALNDTVITSDVSGTVVEVNSDIDPASKTSQVLHVATEG 239

-922-

QIQDLNDAYADAQAEVNKAQ ALNDTV+ S VSGTVVEVN+DIDP+SK SQ LVHVATEG
 Sbjct: 171 QIQDLNDAYADAQAEVNKAQIALNDTVVVISSVSGTVVEVNNDIDPSSKNSQTLVHVATEG 230
 Query: 240 KLQVQGIMSEYDLANVKKDQAVKIKSKVYPDKKEWEGKISYISNYP-EAEANN-----NDS 293
 +LQV+GT++EYDLANVK Q+VKIKSKVY ++EW GKISY+SNYP E+ A + +
 Sbjct: 231 QLQVKGTLTEYDLANVKVGQSVKIKSKVYSNQEWTKGISYVSNYPTESNAGSTTFAGSTG 290
 Query: 294 NNGSSAVNYKYKVDITSPDLALKQGFTVSVEVVNGDKHLIVPTSSVINKDNKHFVWVYND 353
 S+ Y YK+DI SPL+ LKQGFTVSVEVVN K +VP ++VI KD KH+VW Y+D
 Sbjct: 291 AGSSTGATYDYKIDIISPLNQLKQGFTVSVEVVNEAKQALVPLTAVIKKDKKHVWVYD 350
 Query: 354 SNRKISKVEVKIGKADAKTQEILSGLKAGQIVVTNPSKTFKDGQKIDNIESIDLNSNKKSE 414
 + K KVEV +G ADA+ QEI G+ G IV+ NP K K +K++ + SI N+ + +
 Sbjct: 351 ATGKAKKVEVTLGNADAQQQEIHKGVAVGDIVIANPDKNIPDKKLEGVISIGTNTKPEKD 411

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 839

A DNA sequence (GBSx0891) was identified in *S.agalactiae* <SEQ ID 2535> which encodes the amino acid sequence <SEQ ID 2536>. Analysis of this protein sequence reveals the following:

Possible site: 29
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1832(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 840

A DNA sequence (GBSx0892) was identified in *S.agalactiae* <SEQ ID 2537> which encodes the amino acid sequence <SEQ ID 2538>. This protein is predicted to be carbamoyl-phosphate synthase, pyrimidine-specific, large chain, putati. Analysis of this protein sequence reveals the following:

Possible site: 59
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -1.70 Transmembrane 486 - 502 (486 - 502)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.1680(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA91005 GB:Z54240 carbamoyl-phosphate synthase [Lactobacillus
 plantarum]
 Identities = 117/417 (28%), Positives = 205/417 (49%), Gaps = 37/417 (8%)
 Query: 122 FVQVDCLVMRDSLNNCLYVSDLEYIES-NKTTGKSLAIVPSQTLSDAARQTIRDVAFDVC 180
 + +++ VMRD+ +N + V ++E + TG S+ P QTL+D Q +RD A +
 Sbjct: 213 YKEIEFEVMRDAADNAMVVCNMENFDPVGIHTGDSIVYAPVQTLADREVQLLRDAALKII 272

-923-

Query: 181 RKANIIGVCYFSFLIDLNSLDYHIISSGLSHQSILFETITITYPVLEIATKLTGVTFS 240
 R I G C +D NS +Y+II ++ +S S L T YP+ ++A K+ VG
 Sbjct: 273 RALKIEGGCNVQLALDPNSFNYYIIIEVNPVRSSALASKATGYPIAKMAAKIAVGLHLD 332

5 Query: 241 QLKHSYYPNTSAFLEPQLDYVATV--SFSFEKVDY-----IFFARNIEQL 283
 ++K+ T A EP LDYV + F+K + + RNIE+
 Sbjct: 333 EIKNFVTGTTTAAEFEPALDYVVKIPRWPFDPKFTHADRLGTQMKAEGVMAIGRNIEEA 392

10 Query: 284 FTLNLEASS---HDHFFFLSDISEEDLMFALIQKKENRLAYLLEAFRRGFDLYDLSSVT 339
 L + + H L + ++ L LI +++RL YL EA RRG+ + +L+ +T
 Sbjct: 393 TLKAVRSLEIGVHHVEESTLRSVDDVLSDKLIHAQDRLFYLTEAIRRGYQIDELAELE 452

15 Query: 340 KINPFYLDKCLHIVELYENLNKSQYNVDIYKEAKRYGFSDDYIASSWQISLIDMLEYRKK 399
 KIN F+LDK LHI+E+ + L +++ AKR GF+D +A W ++ + ++R
 Sbjct: 453 KINVFFLDKLLHIIIEIEQALRTHTDIETLTVAKRNGFADQTVADYWHETIDQVRDFRLA 512

20 Query: 400 HSVAPVLKQVEQSSGVLTHGHQIQYFRSYDWHSDYISSGCQKALIM-----VDKGY 449
 H +APV K V+ +G Y+ +Y++ ++ I + L++ V+ Y
 Sbjct: 513 HKLAPVYKMWDTTCAGEFASETPYYYGTYEFENESIVTKRPSVLVLGSGPIRIGQGVDFDY 572

Query: 450 SLVKLNELIKQIKQTHLELLIVTNQPLLEQLNDTS--IIFDTIGIETILTIMGIEE 504
 + V +K I++ E +I+ + P + S + F+ + IE +L ++ +E+
 Sbjct: 573 ATV---HSVKAIQKAGYEAIIIMNSNPETVSTDFSVSDKLYFEELTIEDVLNVIELEK 626

25 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 841

A DNA sequence (GBSx0893) was identified in *S.agalactiae* <SEQ ID 2539> which encodes the amino acid sequence <SEQ ID 2540>. This protein is predicted to be carbamoyl phosphate synthetase small subunit (carA). Analysis of this protein sequence reveals the following:

Possible site: 27

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

35 bacterial cytoplasm --- Certainty=0.2709(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

40 >GP:CAB89872 GB:AJ132624 carbamoyl phosphate synthetase small
 subunit [Lactococcus lactis]
 Identities = 188/352 (53%), Positives = 265/352 (74%)

45 Query: 1 MAKKLLILEDGTVFEGLSFGSSLDVTGELVFCTGNTGYQEIITNPSHNGKILVFTSPLIG 60
 M+K+LLILEDGT+FE+G + G++LDVTGELVF TG TGYQE IT+ S+NG+IL FT P++G
 Sbjct: 1 MSKRLILEDGTIFEALGANLDVTGELVFNTGMTGYQESITDQSYNGQILTPTYPVIG 60

Query: 61 NYGIHRSYSEAIIPTCIGVVVAEYSRCVSSDTSKMNLEFLKMKKVPAMSGVDTRYLMQV 120
 NYG++R E+I PTC VVV E +R S+ +M+ DEFLK K +P ++GVDTR + ++
 50 Sbjct: 61 NYGVNRDDYESIHPTCKAVVVHEAARPSNWRMQMSFDEFLKSKNIPGITGVDTTRAITKI 120

Query: 121 IKEKGVFKATLAEAGDVLSHLQDQLIATVLPNTNVEQVSTKTAYPSPASGRNIVLDFGL 180
 ++E G +KA+L +A D + H QL ATVLPNTN VE ST TAYPSP +GR +VV+DFGL
 55 Sbjct: 121 VREHGTMKASLVQARDEVDHQMSQLQATVLPNTNQVETSSATAYPSPNTGRKVVVDFGL 180

Query: 181 KHSILRELSKRQCDVTVIPYNTSLEGIKNLYPEGIILSNGPGNPEKLQEIILNTIKELQKS 240
 KHSILRELSKR+C++TV+PYNTS + I + P+G++L+NGPG+P + E + IKE+Q
 Sbjct: 181 KHSILRELSKRECNLTVPYNTSAKEILEMEPDGVMLTNGPGDPTDVPEAIEMIKEVQKG 240

60 Query: 241 VEMLGIGLGHQLIAMANGAEIMRLPVAKKGPYPMRDIATGRLETVSQFNHFTVNRNLNP 300
 +P+ GI LGHQL ++ANGA ++ +G N+ +R++ATGR++ SQ + + V+ NLP

-924-

Sbjct: 241 IPIFGICLGHQLFSLANGATTYKMKFGHRGFNHAVREVATGRIDFTSQNHGYAVSSENLP 300

Query: 301 HDLLVTHEGLNDQEIVALRHRSFPVMSVQFYPEAAPGPHDVTYFFDEFLEMI 352

DL++TH +ND + +RH+ FP SVQF+P+AAPGPHD +Y FD+F++++

5 Sbjct: 301 EDLMITHVEINDNSVEGVRHKYFPAPFSVQFHPDAAPGPHDASYLFDDFMDLM 352

There is also homology to SEQ ID 2030.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

10 Example 842

A DNA sequence (GBSx0894) was identified in *S.agalactiae* <SEQ ID 2541> which encodes the amino acid sequence <SEQ ID 2542>. Analysis of this protein sequence reveals the following:

Possible site: 57

>>> Seems to have no N-terminal signal sequence

15

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3646(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

20

A related GBS nucleic acid sequence <SEQ ID 9967> which encodes amino acid sequence <SEQ ID 9968> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

25 >GP:CAB89869 GB:AJ132624 pyrimidine regulatory protein [Lactococcus

lactis]

Identities = 127/169 (75%), Positives = 147/169 (86%)

Query: 13 MKRKEIIDVMTKRAITRITYEIIERNKNLDNIVLAGIKTRGVFLAKRIQERLQLENLD 72

M RKEIID++TMKRAITRITYEIIERNK LD +VL GIKTRGV+LAKRIQERL+QLE L+

30

Sbjct: 1 MARKEIIDBITMKRAITRITYEIIERNKELDKLVILIGIKTRGVYLAKRIQERLQLEGLE 60

Query: 73 IPVGEIDTKPFRDDMKVEVDITTMPVDITDKDIILIDDVLYTGRTIRAAIDNLVSLGRPS 132

IP GEIDT+PFRDD + + DIT + +DIT KD+IL+DDVLYTGRTIRAAID +V LGRP+

35

Sbjct: 61 IPFGEIDTRPFRDDKQAQEDTTEIDIDITGKDVLVDDVLYTGRTIRAAIDGIVKLGRPA 120

Query: 133 RVSLAVLIDRGHRELPIRADYVGKNIPTSQFEEILVEVMEHDGYDRVSI 181

RV LAVL+DRGHRELPIRADYVGKNIPT EEI+V++ EHDG D + I

Sbjct: 121 RVQLAVLVDRGHRELPIRADYVGKNIPTGHDEETIVQMSEHDGNSILI 169

40 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2543> which encodes the amino acid sequence <SEQ ID 2544>. Analysis of this protein sequence reveals the following:

Possible site: 45

>>> Seems to have no N-terminal signal sequence

45

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3870(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

50 An alignment of the GAS and GBS proteins is shown below.

Identities = 147/171 (85%), Positives = 158/171 (91%)

Query: 13 MKRKEIIDVMTKRAITRITYEIIERNKNLDNIVLAGIKTRGVFLAKRIQERLQLENLD 72

MK KEI+DDVMTKRAITRITYEIIERNK LDN+VLGKTRGVFLA+RIQERL QLE LD

55

Sbjct: 1 MKTKRIVDDVMTKRAITRITYEIIERNKQLDNVVLGKTRGVFLARRIQERLHQLEGLD 60

-925-

Query: 73 IPVGELDTKPFRRDDMKVEVDTTMPVDITDKDIILIDDVLYTGRTTIRAAIDNLVSLGRPS 132
 +P+GELD KPFRRDM+VE DTT M VDIT KD+ILIDDVLYTGRTTIRAAIDNLVSLGRP+
 Sbjct: 61 LPIGELDIKPFRRDMRVEEDTTLMSSVDITGKDVILIDDVLYTGRTTIRAAIDNLVSLGRPA 120

5 Query: 133 RVSLAVLIDRGHRELPIRADYVGKNIPTSQFEEILVEVMEHDGYDRVSIID 183
 RVSLAVL+DRGHRELPIRADYVGKNIPTS EEI+VEV+E DG DRVSIID
 Sbjct: 121 RVSLAVLVDRGHRELPIRADYVGKNIPTSSVEEIVVEVVEVDGRDRVSIID 171

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 10 vaccines or diagnostics.

Example 843

A DNA sequence (GBSx0895) was identified in *S.agalactiae* <SEQ ID 2545> which encodes the amino acid sequence <SEQ ID 2546> (rluD). Analysis of this protein sequence reveals the following:

Possible site: 35
 15 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0687(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 20 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9969> which encodes amino acid sequence <SEQ ID 9970> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

25 >GP:BAB06261 GB:AP001515 unknown conserved protein [Bacillus halodurans]
 Identities = 178/290 (61%), Positives = 216/290 (74%), Gaps = 2/290 (0%)

Query: 17 GVRLDKAL-ADNSELRSQANEEIKKGIVLVNGQVKKAKYTVQEGDRITFDIPKEEVLVDY 75
 G R+DK L A E SR+Q + IK G VL+NG+ K+ Y V+ GD + +P+ EVL+
 30 Sbjct: 15 GERIDKFLTAQGEWSRTQVQWIKDGHVLLINGRTIKSNYKVETGDTLELFVPEPEVLEV 74

Query: 76 QAENIPLDIIYQDDDDVAVVNKPQGMVVHPSAGHSSGTLVNALMYHIKDLSSINGVVRPGI 135
 ENIP++IIY+D+DVAVVNKP+GMVVHP+ GH++GTLVNALMYH DLSSINGVVRPGI
 35 Sbjct: 75 VPENIPIEIIYEDEDVAVVNKPGRGMVVHPAPGHTTGTLVNALMYHCNDLSSINGVVRPGI 134

Query: 136 VHRIDKDTSGLLMVAKNDRAHQVLAELKDKKSLRKYLAIHVGNLNPDRGVIEAPIGRSD 195
 VHRIDKDTSGLLM+AKNDRAH+ L +LK K + R Y AIVHGN+P+D G I+APIGR
 40 Sbjct: 135 VHRIDKDTSGLLMIAKNDRAHESLVNQLKAKTTERVYQAIHVGNIPDHGTTIDAPIGRDK 194

Query: 196 KDRKKQAVTAK-GKEPAITRFHVLERFGDYTLVELSLETGRTHQIRVHMAYIGHPLAGDPV 254
 DR+ VT + + A+T F VLERFGD+T VE LETGRTHQIRVH YIG PLAGDP
 45 Sbjct: 195 VDRQSMVTVEENSRAVDVHTFTVLERFGDFTFVEQLETCGRTHQIRVHPKYIGFPLAGDPK 254

Query: 255 YGPRKTLGGKGQFLHAQTLGFTHPNNGENLIFSVEVPEIFQTTLKLRKN 304
 YGP+KTL GQ LHAQ LGF HP GE + F VE+PE + +L+ N
 50 Sbjct: 255 YGPKKTLSDGQALHAQKLGFEHPRTGEFMRFKVEMPEEMKKLIRQLQNN 304

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2547> which encodes the amino acid sequence <SEQ ID 2548>. Analysis of this protein sequence reveals the following:

50 Possible site: 27
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2455(Affirmative) < succ>
 55 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

-926-

Identities = 239/295 (81%), Positives = 265/295 (89%)

Query: 9 MEITIKIAGVRLDKALADNSELRSQANEEIKKGIVLVNGQVKAKYTVQEGDRITFDIP 68
 MEI + +G RLDKALAD S LSR QAN++IK+G+VLVNGQ KKAKYTVQ GD I F++P
 5 Sbjct: 1 MEINVITSGQRLDKALADLSPLSRGQANDQIKQGLVLVNGQQKKAKYTVQAGDVICFELP 60

Query: 69 KEEVL DYQAENIPLDIITYQDDDAVAVNKPQGMVVHPSAGHSSGTLVNALMYHIKDLSSIN 128
 KEEVL+YQA+NIPLDIITY+DD +A++NKPQGMVVHPSAGH SGT+VNALMYHIKDLSSIN
 10 Sbjct: 61 KEEVLEYQAQNIPLDIITYEDDALAIINKPQGMVVHPSAGHPSGTMVNALMYHIKDLSSIN 120

Query: 129 GVVVRPGIVHRIDKDTSGLLMVAKNDRHQVLABELKDKKSLRKYLAIVHGNLPNDRGVIE 188
 GVVVRPGIVHRIDKDTSGLLMVAK D AHQ LABELK KSLRKYLAIVHGNLPNDRG+IE
 15 Sbjct: 121 GVVVRPGIVHRIDKDTSGLLMVAKTDAAHQALABELKAKKSLRKYLAIVHGNLPNDRGMIE 180

Query: 189 APIGRSDKDRKKQAVTAKGKPAIRFHVLERFGDYTLVELSLETGRTHQIRVHMAYIGHP 248
 APIGRS+KDRKKQAVTAKGK A+TRF VLERFGDY+LVEL LETGRTHQIRVHMAYIGHP
 20 Sbjct: 181 APIGRSEKDRKKQAVTAKGKEAVTRFTVLERFGDYSVELQLLETGRTHQIRVHMAYIGHP 240

Query: 249 LAGDPVYGPBKTLGGKGQFLHAQTLGFTHPNNGENLIFSVEVPEIFQTTLKLRK 303
 +AGDP+YGPBKTL G GQFLHA+TLG THP G+ +IF+VE PEIFQ L+ LRK
 25 Sbjct: 241 VAGDPVYGPBKTLSGHGQFLHAKTLGLTHPMTGKEMIFTVEAPEIFQKVLKLLRK 295

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

25 Example 844

A DNA sequence (GBSx0896) was identified in *S.agalactiae* <SEQ ID 2549> which encodes the amino acid sequence <SEQ ID 2550>. Analysis of this protein sequence reveals the following:

Possible site: 33

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0496 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP: AAD53064 GB: AF163833 CpsY [Streptococcus agalactiae]

Identities = 105/297 (35%), Positives = 163/297 (54%), Gaps = 4/297 (1%)

Query: 1 MNIQQLRYVVAIANSQTFREAAAKLFVSQPSLSVAVRDLETELGFIQIFTRTTTGAVLTNQ 60
 M IQQL+YV+ I +G+ EAA +L+++QPSLS AVR+LETE+G QIF R G LT
 40 Sbjct: 1 MRIQQLQYVIKIVETGSMNEAAKQLYITQPSLSNAVRNLETEMGIQIFIRNPKGITLTKD 60

Query: 61 GMTFYENALEVVKSFDSEKQFSQSEATEQBFASQHYDFLPPLITAFSKCNDNFYSY-F 119
 GM F A ++++ E+++ + + FS++SQHY F+ A D Y
 45 Sbjct: 61 GMEFLSYARQILEQTALLEERYKGDNTSRELFVSQHYAFVVNAFVALFNGIDMTQYEL 120

Query: 120 RIFESTTIRILDEVAQGNSEIGIYYINSQNKGLLQRLDKLGFVELIPFKTHIYLGKD 179
 + E+ T I+D+V SEIG++++NS N+ L + D L L HI++ K
 50 Sbjct: 121 FLRETRTWEIIDDVKNFRSEIGVLFNSYNRDVLTKLFDNLSLIATTLFTTTPHIFVSKS 180

Query: 180 HPLASKTSLIMTDLEGLPTVRFTQDRDDYRYSENFEVLDSSVTYNVTD RATLNGILER 239
 +PLA++ L M DLE P + + Q + Y+SE + + + V+DRATL ++
 55 Sbjct: 181 NPLANRKKLSMKDLEDYPYLSYDQGLHNSFYFSEEMMSQIPHPSIVVSDRATLFLNLMIG 240

Query: 240 TQAYATGSGFLDSRVNG--ITVIPLDHLDNQMIYIKRDKRNLSQLKFKFVAVMEE 294
 Y +G L+S+ +NG I IPL+ ++YI+ NLS+M KF+ + E
 60 Sbjct: 241 LDGYTVATGILNSK-LNGDEIVAIPLDVDDVIDIVYIRHDKANLSKMGQKFIDYILLE 296

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2551> which encodes the amino acid sequence <SEQ ID 2552>. Analysis of this protein sequence reveals the following:

-927-

Possible site: 13

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.1252(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

10 Identities = 217/296 (73%), Positives = 253/296 (85%)

Query: 1 MNIQQLRYVVAIANSGTFREAAAKLFVSPSLSVAVRDLETGLGFQIFTRTTTGAVLTNQ 60
 MNIQQLRYVVAIAN+GTFREAA+KLFVSPSLSV++DLE ELGFQIF RTT+G VLT+Q
 Sbjct: 1 MNIQQLRYVVAIANNGTFREAAKLFVSPSLSVSIKDLEAGLGFQIFNRTTSGTVLTSQ 60

15 Query: 61 GMTTFYENALEVVKSFDSEKQFSQSEATEQEFSIASQHYDFLPPLITAFSKCNDNFSYFR 120
 G+ FYE ALEVVKSFDSFEK FSQ++ + EFSIASQHYDFLPPLITAFS+ D FR
 Sbjct: 61 GLVIFYEKALEVVKSFDSEKTFSQADLDQNEFSIASQHYDFLPPLITAFSQQYDGHVRVFR 120

20 Query: 121 IFESTTIRILDEVAQGNSEIGIYYINSQNKGLLQRLDKLGLFVELIPFKTHIYLGKDH 180
 IFESTTI+ILDEVAQGNSEIGIYY+N N+KGL QR+DKLGLF+V LIFF THIYL K H
 Sbjct: 121 IFESTTIQILDEVAQGNSEIGIYYLNVDNQKGLFQRMKLGLEYVSLIPFTTHIYLSKTH 180

25 Query: 181 PLASKTSLIMTDLEGLPTVRFTQDRDDYRYSENFEVLDSSVTYNVTD RATLNGILERT 240
 PLA++ +L + D++GLP VRFTQ+RD+Y YYSENFV+ + YNNV+DRATLNGILERT
 Sbjct: 181 PLANREALYLNDIQGLPAVRFTQERDEYLYYSENFVDTSECPRIYNNVSDRATLNGILERT 240

30 Query: 241 QAYATGSGFLDSRSVNGITVIPLDHLNQMIYIKRKDRNLSQMAKLFVAVMEEF 296
 A+ATGSGFLD RSVNGI VIPL DH+DNQMIY+KRKD+NLS FV ++++YF
 Sbjct: 241 NAFATGSGFLDHRSVNGIKVIPLADHIDNQMIYVKKDKNLSVAGATFVTILKDYF 296

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 845

35 A DNA sequence (GBSx0897) was identified in *S.galactiae* <SEQ ID 2553> which encodes the amino acid sequence <SEQ ID 2554>. This protein is predicted to be 50S ribosomal protein L27 (rpmA). Analysis of this protein sequence reveals the following:

Possible site: 36

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

40 bacterial cytoplasm --- Certainty=0.0976(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB14754 GB:Z99118 ribosomal protein L27 (BL24) [Bacillus subtilis]
 Identities = 70/90 (77%), Positives = 80/90 (88%)

50 Query: 8 NLQLFAHKKGGGSTSNGRDSQAKRLGAKAADGQTVSGGSILYRQRTGTHIYPGANVGRGGD 67
 +LQ FA KKG GST NGRDS+AKRLGAK ADGQ V+GGSILYRQRT IYPG NVGRGGD
 Sbjct: 5 DLQFFASKKGVGSTKNGRDSEAKRLGAKRADGQFVTGGSILYRQRTKIYPGENVGRGGD 64

55 Query: 68 DTLFAKVEGVVRFERKGRDKKQVSVYPIAK 97
 DTLFAK++G V+FER GRD+K+VSVYP+A+
 Sbjct: 65 DTLFAKIDGTVKFERFGRDRKKVSVYPVAQ 94

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2555> which encodes the amino acid sequence <SEQ ID 2556>. Analysis of this protein sequence reveals the following:

-928-

Possible site: 36
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.0976(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

10 Identities = 95/97 (97%), Positives = 96/97 (98%)

Query: 1 MLKMNLANLQLFAHKKGGGSTSNGRDSQAKRLGAKAADGQTVSGGSILYRQRGTHIYPGA 60
 MLKMNLANLQLFAHKKGGGSTSNGRDSQAKRLGAKAADGQTVSGGSILYRQRGTHIYPG

15 Sbjct: 1 MLKMNLANLQLFAHKKGGGSTSNGRDSQAKRLGAKAADGQTVSGGSILYRQRGTHIYPGV 60

Query: 61 NVGRGGDDTLFAKVEGVVRFERKGRDKKQVSVYPIAK 97

 NVGRGGDDTLFAKVEGVVRFERKGRDKKQVSVYP+AK

15 Sbjct: 61 NVGRGGDDTLFAKVEGVVRFERKGRDKKQVSVYPVAK 97

20 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 846

A DNA sequence (GBSx0898) was identified in *S.agalactiae* <SEQ ID 2557> which encodes the amino acid sequence <SEQ ID 2558>. Analysis of this protein sequence reveals the following:

25 Possible site: 25

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -0.75 Transmembrane 32 - 48 (32 - 48)

----- Final Results -----

30 bacterial membrane --- Certainty=0.1298(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

35 >GP:BAB06729 GB:AP001517 unknown conserved protein in B. subtilis

[Bacillus halodurans]

Identities = 33/107 (30%), Positives = 63/107 (58%), Gaps = 4/107 (3%)

40 Query: 1 MIKATFTRNQSGYLYSAEISCHAGSGEYGFVDICAAVSTLSINFINSLEALTTCQAQLII 60

MI F RN+ + S +SGHA +G YG D++CA S +++ +N++ AL CQ +L+

40 Sbjct: 1 MIDVVFERNKQNDIVSFTMSGHADAGPYQDLVCAGASAVALTGNVAILAL--CQVELVT 58

Query: 61 N-DVEGGYMKIDL-SSIPQHKEDKVQLLFESYLLGMTNLSKDSSEFV 105

+ EGG+++ + + + + +KVQLL E + + ++++ E +

45 Sbjct: 59 EMENEGGFRLCRVPNDLEETTFEKVQLLLEGMNISLQSIASYSYGEHI 105

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2559> which encodes the amino acid sequence <SEQ ID 2560>. Analysis of this protein sequence reveals the following:

Possible site: 52

50 >>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -0.59 Transmembrane 32 - 48 (32 - 48)

----- Final Results -----

55 bacterial membrane --- Certainty=0.1235(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

-929-

>GP:EAB06729 GB:AP001517 unknown conserved protein in *B. subtilis*
 [Bacillus halodurans]
 Identities = 33/109 (30%), Positives = 60/109 (54%), Gaps = 4/109 (3%)

5 Query: 1 MIKAIFTTRQKNGQLSSVTLTGHAGSGKHGFDIVCASVSTLAINFVNSLEVLAQCQALVDL 60
 MI +F R K + S T++GHA +G +G D+VCA S +A+ VN++ L + + ++
 Sbjct: 1 MIDVVFERNKQNDIVSFTMSGHADAGPYGQDLVCAGASAVALGTVNAILALCQVELVTEM 60

10 Query: 61 NDVEGGYMAITIP---PHDNKEEVQLLFESFLLGMTSLAKDSSKFVNTQ 106
 + EGG++ +P E+VQLL E + + S+A+ + + +
 Sbjct: 61 EN-EGGFLRCRVPNDLEETTFEKKVQLLLEGMNISLQSIASVGEHIQIE 108

An alignment of the GAS and GBS proteins is shown below.

Identities = 67/110 (60%), Positives = 90/110 (80%), Gaps = 2/110 (1%)

15 Query: 1 MIKAIFTTRQSGYLYSAEISGHAGSGEYGFDIVCAAVSTLSINFINSLEALTTCQAQLII 60
 MIKA FTR ++G L S ++GHAGSG++GFD++CA+VSTL+INF+NSLE L CQA + +
 Sbjct: 1 MIKAIFTTRQKNGQLSSVTLTGHAGSGKHGFDIVCASVSTLAINFVNSLEVLAQCQALVDL 60

20 Query: 61 NDVEGGYMKIDLSSIPQHKEDKVQLLFESYLLGMTNLKSDSEFVSTVVM 110
 NDVEGGYM I + P +++VQLLFES+LLGMT+L+KDSS+FV+T V+
 Sbjct: 61 NDVEGGYMAITIP---PHDNKEEVQLLFESFLLGMTSLAKDSSKFVNTQVI 108

SEQ ID 2558 (GBS433) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell
 25 extract is shown in Figure 78 (lane 4; MW 16kDa). It was also expressed in *E.coli* as a GST-fusion
 product. SDS-PAGE analysis of total cell extract is shown in Figure 173 (lane 8; MW 41kDa).

GBS433-GST was purified as shown in Figure 223, lane 10.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 vaccines or diagnostics.

30 Example 847

A DNA sequence (GBSx0899) was identified in *S.agalactiae* <SEQ ID 2561> which encodes the amino
 acid sequence <SEQ ID 2562>. This protein is predicted to be ribosomal protein L21 (rplU). Analysis of
 this protein sequence reveals the following:

Possible site: 57
 35 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2972 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 40 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB14756 GB:Z99118 ribosomal protein L21 (BL20) [Bacillus subtilis]
 Identities = 67/101 (66%), Positives = 78/101 (76%)

45 Query: 4 YAIIKTGGKQVKVEVGQAIYVEKLDVEAGAEVTFNEVVLVGGETTKVGTTPVVEGATVVGT 63
 YAIIKTGGKQ+KVE GQ +Y+EKL EAG VTF +V+ VGG+ KVG P VEGATV
 Sbjct: 2 YAIIKTGGKQIKVEEGQTVYIEKLAEEAGETVTFEDVLFVGGDNVKGVNPTVEGATVTAK 61

50 Query: 64 VEKQGGKQKKVVSYYKPKKSGHRQGHROPYTKVVINAINA 104
 VEKQG+ KK+ ++YKPKK H+KQGHROPYTKV I INA
 Sbjct: 62 VEKQGRAKKITVFRYKPKINVHKQGHROPYTKVTIEKINA 102

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2563> which encodes the amino acid
 55 sequence <SEQ ID 2564>. Analysis of this protein sequence reveals the following:

-930-

Possible site: 33

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.3026(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

10 Identities = 97/104 (93%), Positives = 101/104 (96%)

Query: 1 MSTYAIIKTGGKQVKVEVGQAIYVEKLDVEAGAEVTFNEVVLVGGETTKVGTTPVVEGATV 60

MSTYAIIKTGGKQVKVEVGQAIYVEK+D EAGAEVTFNEVVLVG+ T VGTTPVVEGATV

Sbjct: 1 MSTYAIIKTGGKQVKVEVGQAIYVEKIDAEAGAEVTFNEVVLVGDKTVVGTTPVVEGATV 60

15

Query: 61 VGTVEKQKGKQKKVVSYYKPKKGSHRKQHRQPYTKVVINAINA 104

VGTVEKQKGKQKKVV++KYKPKKGSHRKQHRQPYTKVVINAINA

Sbjct: 61 VGTVEKQKGKQKKVVTFKYKPKKGSHRKQHRQPYTKVVINAINA 104

20 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 vaccines or diagnostics.

Example 848

A DNA sequence (GBSx0900) was identified in *S.galactiae* <SEQ ID 2565> which encodes the amino
 acid sequence <SEQ ID 2566>. Analysis of this protein sequence reveals the following:

25 Possible site: 16

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

30 bacterial cytoplasm --- Certainty=0.1032(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9369> which encodes amino acid sequence <SEQ ID 9370>
 was also identified.

35 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB14809 GB:Z99118 excinuclease ABC (subunit C) [Bacillus subtilis]

Identities = 221/373 (59%), Positives = 288/373 (76%)

40 Query: 1 MKSAAMIMEFERAAEYRDLEIAISLLRTKQRVIHQDMKDRDVFVFDKGMCMCVQVFFVR 60
 M AA +EFERA E RD I I KQ++ D+ DRDVF Y DKGWMCVQVFF+R
 Sbjct: 206 MHEAAENLEFERAKELRDQIAHIESTMEKQKMTMNDLVDRDVFAYAYDKGMCMCVQVFFIR 265

45 Query: 61 NGKLIQRDVMNMFPPYNEFEEDFLTYIGQFYQDTKHFLPKKEVFIPQDIDAKSVETIVGCKI 120
 GKLI+RDV+MFP Y E +E+FLT+IGQFY HFLPKE+ +P ID +E ++ +
 Sbjct: 266 QGKLIERDVSMFPLYQEADDEFLTFIGQFYSKNNHFLPKEILVPDSIDQSMIEQLLETNV 325

50 Query: 121 VKPQRGEKKQLVNLAIKARVSLQKFDLLEKDIRKTHGAIENLGNLNIKPKPVRIEAFD 180
 +P++G KK+L+ LA KNA+++L++KF L+E+D ++ GA++ LG LNI P RI AFD
 Sbjct: 326 HQPKKGPKKELMLAHKNAKIALKEKFSLLIERDEERSIGAVQKLGEALNIYTPHRIVAFD 385

55 Query: 181 NSNIQGTSPVAMVVFVNGKPSKKDYRKFKIKTVIGPDDYASMREVIHRRYSRVLKDLGT 240
 NSNIQGT+PVA+AM+VF++GKP KK+YRK+KIKTV GPDDY SMREV+ RRY+RVL++ L
 Sbjct: 386 NSNIQGTNPVSAMIVFIDGKPKYKRYKIKTVTGPDDYGSMSREVRRRYTRVLRENLP 445

55 Query: 241 PPDLIVIDGGQGVNIARDVIENQFGLAIPAGLQKNDKHQTHELLFGDPLEVVELPRNS 300
 PDLI+IDGG+GQ+N ARDVIE+ GL IPIAGL K++KH+T LL GDPLEV L RNS
 Sbjct: 446 LPDLIIDGGKGQINAARDVIEENELGLDPIAGLAKDEKHRTSNLLIGDPLEVAYLERN 505

Query: 301 EEEFFLLHRIQDEVHRFAITFHRQLRSKNSFSKLDGITGLGPKRKQLLMKHFSLPNIQK 360

-931-

+EF+LL RIQDEVHRFAI+FHRQ+R K++F S LD I G+G KRK++L+KHF S+ +++
 Sbjct: 506 QEFYLLQRIQDEVHRFAISFHRQIRGKSAFQSVLDDIPGIGEKRRKMLLKHFSGSVKKMKE 565

Query: 361 AEIEDIIMCGIPR 373
 A +EDI G+P+
 Sbjct: 566 ASLEDIKKAGVPQ 578

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2567> which encodes the amino acid sequence <SEQ ID 2568>. Analysis of this protein sequence reveals the following:

Possible site: 53
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4332(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 289/385 (75%), Positives = 334/385 (86%)

Query: 1 MKSAAMTMEFERAAEYRDLIEAISLLRTKQRVIHQDMKDRDVFQYFVDKGWMCVQVFFVR 60
 M +A+ M FERAAYRDLI I+ +RTKQRV+ +D++DRD+FGY+VDKGWMCVQVFFVR
 Sbjct: 206 MLAASKBMAFERAAEYRDLISGIATMRKQRVMSKDLQDRDIFGYVVDKGWMCVQVFFVR 265

Query: 61 NGKLIQRDVNMFPYYNEPEEDFLTIVIGQFYQDTKHFLPKVFIPQDIDAKSVETIVGCKI 120
 GKLIQRDVN+FPYY + EEDFLTIV+GQFYQD +HF+PKEVFIP+ ID + V IV KI
 Sbjct: 266 QGKLIQRDVNLFPPYYTDAEEDFLTVMGQFYQDKQHFIPKEVFIPKAIDEELVAAIVPTKI 325

Query: 121 VKPQRGEKKQLVNLAIKNARVSLQQKFDLLEKDIRKTHGAIVENLGNLNPVRIEAFD 180
 +KP+RGEKKQLV LA KNARVSLQQKFDLLEKDI+KT GAIVENLG LL I KPVRIEAFD
 Sbjct: 326 IKPKRGEKKQLVALATKNARVSLQQKFDLLEKDIKKTGAIVENLGQLLRIDKPVRIEAFD 385

Query: 181 NSNIQGTSPVAAMVVFVNGKPSKKDYRKFKIKTVIGPDDYASMREVIHRRYSRVLKDGLT 240
 NSNIQGTSPVAAMVVFV+GKPSKKDYRKFKIKTV+GPDDYASMREV+ RRYSRV K+GL
 Sbjct: 386 NSNIQGTSPVAAMVVFVDGKPSKKDYRKFKIKTVVGPDDYASMREVLFRYSRVKKEGLQ 445

Query: 241 PPDLLIVDGGQGVNIARDVIENQFGLAIPAGLQKNDKHQTHELLFGDPLEVVELPRNS 300
 P+LI++DGG GQVN+A+DVIE Q GL IP+AGLQKNDKHQTH+LLFG+PLEVV LPR S
 Sbjct: 446 APNLIIVDGGVGVQVNVAKDVIEKQLGLTIPVAGLQKNDKHQTHDLLFGNPLEVVELPRRS 505

Query: 301 EEFFLLHRIQDEVHRFAITFHRQLRSKNSFSKLDGITGLGPKRKQLLMKHFKSLFNIQK 360
 EEFFLLHRIQDEVHRFA+TFHRQ+R KNSFSS LD I+GLGPKRKQLL++HFK++ I
 Sbjct: 506 EEFFLLHRIQDEVHRFAITFHRQVRRKNSFSSTLDHISGLGPKRKQLLLRHFKTTITAS 565

Query: 361 AEIEDIIMCGIPRTVAESLRDSIND 385
 A E+I GIP+TV E+++ + D
 Sbjct: 566 ATSEEIQALGIPKTVVEAIQQQITD 590

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 849

A DNA sequence (GBSx0901) was identified in *S.agalactiae* <SEQ ID 2569> which encodes the amino acid sequence <SEQ ID 2570>. Analysis of this protein sequence reveals the following:

Possible site: 16
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2491(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 850

A DNA sequence (GBSx0902) was identified in *S.agalactiae* <SEQ ID 2571> which encodes the amino acid sequence <SEQ ID 2572>. Analysis of this protein sequence reveals the following:

```

Possible site: 55
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3349(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:BAA86651 GB:AB033763 glycerophosphoryl diester phosphodiesterase
      homologue [Staphylococcus aureus]
Identities = 50/202 (24%), Positives = 96/202 (46%), Gaps = 15/202 (7%)

Query: 1  MDVIMTKDHKLVIHDDNLKRLSGMKNKDVSKLTLDQVTKIPIHQ---GRFA-SHIPSFTE 56
      +DV +TKD +L++IHDD L+R + M+ ++++L D++ +F H+P+F +
Sbjct: 36  LDVAITKDEQLIIHDDYLERITNMSGEITELNYDEIKDASAGSWFGEKFKDEHLPTFDD 95

Query: 57  FMKTAQSLDQKIMIELKPY-NQNLDIYADEFIKEFK-----LRLSTKHKVMSLNLTLEIK 111
      +K A + + +ELK N + +K+ +E L + + + S N+ L++
Sbjct: 96  VVKIANEYNNMLNLVELKGITGPNGLALSKSMVKQVEEQLTNLNQNEVLISFNVVLVKL 155

Query: 112 VEKKLPQLDTGYLIPL-----HWGTLQNH-NVDFYGIIEFSYNDWIAYLAQEYNKQLYVW 165
      E+ +PQ + + W TL ++ N E+ + +E +L VW
Sbjct: 156 ABEIMPQYNRAVIFHTTSFREDWRTLDDYCNAKIVNTEDAKLTAKVKMVKKEAGYELNVW 215

Query: 166 TINRDNLMI RYLQSPVNGIITD 187
      T+N+ + V+GI TD
Sbjct: 216 TVNKPARANQLANWGVGIFTD 237

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2573> which encodes the amino acid sequence <SEQ ID 2574>. Analysis of this protein sequence reveals the following:

```

Possible site: 36
>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -12.26 Transmembrane 239 - 255 ( 227 - 260)
INTEGRAL Likelihood = -9.45 Transmembrane 80 - 96 ( 78 - 108)
INTEGRAL Likelihood = -9.13 Transmembrane 137 - 153 ( 131 - 160)
INTEGRAL Likelihood = -4.94 Transmembrane 278 - 294 ( 277 - 295)
INTEGRAL Likelihood = -3.56 Transmembrane 36 - 52 ( 33 - 55)
INTEGRAL Likelihood = -3.56 Transmembrane 188 - 204 ( 185 - 206)
INTEGRAL Likelihood = -3.35 Transmembrane 314 - 330 ( 310 - 331)

----- Final Results -----
      bacterial membrane --- Certainty=0.5904(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

>GP:CAB12801 GB:Z99109 similar to glycerophosphodiester

```

-933-

phosphodiesterase [Bacillus subtilis]
 Identities = 67/244 (27%), Positives = 110/244 (44%), Gaps = 14/244 (5%)

5 Query: 344 VIAHRLVSAGVENSLEALEGAKKAGSDYVELDLILTNDHFVSHDNRLKRLAGVNTII 403
 +IAHRG EN++ A + A K +D +ELD+ LTKD VV HD+R+ R + +
 Sbjct: 3 IIAHRGASGYAPENTIAAFDLAVKMNADMIELDVQLTKDRQIVVIHDDRVDRTTNGSGFV 62

10 Query: 404 RNLTLKEVEHLTSHQGH---FSGRFVSFDTFYQKAKKLNMPLLIELKPIGTEPGNYVDLF 460
 ++ TL+E++ L + + F G + K + LLIELK ++ G ++
 Sbjct: 63 KDFTLEELQKLDAGSWYGPAPCGERIPTEAVLKRYHKKIGLLIELKGHP SQVGIEEEVG 122

15 Query: 461 LETYHRLGISKDNKVMSLDLEVIEAIIKKKNPSITTGYYIPIQFGFFG-----DEFVDF 513
 + + S +N V S ++ ++ PSI T I FG F ++
 Sbjct: 123 -QLLGQFSFSINNIVQSFGQFSVQRFRELYPSIPTAVITRPNFGMLSRNQMKAFRSFANY 181

20 Query: 514 YVIEDFSYRSYLSSQAFWNKEIYVWTINDPKRIEHYLLKPIQGIITDQPALTNQLIKDL 573
 I+ + N I+ WT+N+ K + GI+TD P + +IKD
 Sbjct: 182 VNKHTRLNRLMIGSINKNGLNIFAWTVNNQKTAAKLQAMGVDGIVTDYP---DFIIKDG 238

Query: 574 KQDN 577
 K +N
 Sbjct: 239 KHEN 242

An alignment of the GAS and GBS proteins is shown below.

25 Identities = 90/215 (41%), Positives = 136/215 (62%)

Query: 1 MDVIMTKDHKLVIHDDNLKRLSGMKNKDVSKLTLDQVTKIPIHQGRFASHIPSFTEFMKT 60
 +D+I+TKD+ VV HD+ LKRL+G+NK + LTL +V + HQG F+ SF F +
 Sbjct: 375 LDLILTNDHFVSHDNRLKRLAGVNTIRNLTLEKEVEHLTSHQGHFSGRFVSFDTFYQK 434

30 Query: 61 AQSLDQKIMIELKPYNQNLDIYADEFIKEFKELRLSTKHKVMSLNLTLEKVEKKLPQLD 120
 A+ L+ ++IELKP Y D F++ + L +S +KVMSL+L +IE ++KK P +
 Sbjct: 435 AKKLNMPLLIELKPIGTEPGNYVDLFLETYHRLGISKDNKVMSLDLEVIEAIIKKKNPSIT 494

35 Query: 121 TGYLIPLHWGTLQNHNVDFYGIIEEFSYNDWIAYLAQYKNQLYVWTINRDNLMIRYLQSP 180
 TGY+IP+ +G + VDFY IE+FSY +++ A NK++YVWTIN + YL P
 Sbjct: 495 TGYIPIQFGFFGDEFVDFYVIEDFSYRSYLSSQAFWNKEIYVWTINDPKRIEHYLLKP 554

40 Query: 181 VNGIITDELNLFKVINIKDKNPNYYQALQLIDS 215
 + GIITD+ L + KD+K +Y+ R +++I S
 Sbjct: 555 IQGIITDQPALTNQLIKDLKQDNSYFSRLVRIISS 589

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

45 Example 851

A DNA sequence (GBSx0903) was identified in *S. agalactiae* <SEQ ID 2575> which encodes the amino acid sequence <SEQ ID 2576>. Analysis of this protein sequence reveals the following:

Possible site: 53
 >>> Seems to have no N-terminal signal sequence

50 INTEGRAL Likelihood = -15.02 Transmembrane 84 - 100 (76 - 112)
 INTEGRAL Likelihood = -3.50 Transmembrane 139 - 155 (139 - 157)
 INTEGRAL Likelihood = -2.23 Transmembrane 41 - 57 (39 - 59)
 INTEGRAL Likelihood = -0.96 Transmembrane 179 - 195 (179 - 195)

55 ----- Final Results -----
 bacterial membrane --- Certainty=0.7007(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

60 A related GBS nucleic acid sequence <SEQ ID 9901> which encodes amino acid sequence <SEQ ID 9902> was also identified.

-934-

The protein has no significant homology with any sequences in the GENPEPT database, but there is homology to SEQ ID 2574.

A related GBS gene <SEQ ID 8671> and protein <SEQ ID 8672> were also identified. Analysis of this protein sequence reveals the following:

```

5      Lipop: Possible site: -1   Crend: 10
      McG: Discrim Score:      -3.38
      GvH: Signal Score (-7.5): -4.08
          Possible site: 53
      >>> Seems to have no N-terminal signal sequence
10     ALOM program   count: 4 value: -15.02 threshold: 0.0
          INTEGRAL   Likelihood = -15.02   Transmembrane   84 - 100 ( 76 - 112)
          INTEGRAL   Likelihood = -3.50    Transmembrane   139 - 155 ( 139 - 157)
          INTEGRAL   Likelihood = -2.23    Transmembrane   41 - 57 ( 39 - 59)
          INTEGRAL   Likelihood = -0.96    Transmembrane   179 - 195 ( 179 - 195)
15     PERIPHERAL   Likelihood = 2.01      104
          modified ALOM score: 3.50

      *** Reasoning Step: 3

20     ----- Final Results -----
          bacterial membrane --- Certainty=0.7007(Affirmative) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 852

A DNA sequence (GBSx0904) was identified in *S.agalactiae* <SEQ ID 2577> which encodes the amino acid sequence <SEQ ID 2578>. Analysis of this protein sequence reveals the following:

```

30     Possible site: 61
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
          bacterial cytoplasm --- Certainty=0.4150(Affirmative) < succ>
35     bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 853

A DNA sequence (GBSx0905) was identified in *S.agalactiae* <SEQ ID 2579> which encodes the amino acid sequence <SEQ ID 2580>. Analysis of this protein sequence reveals the following:

```

45     Possible site: 13
      >>> Seems to have no N-terminal signal sequence
          INTEGRAL   Likelihood = -0.32   Transmembrane   2 - 18 ( 2 - 18)

      ----- Final Results -----
50     bacterial membrane --- Certainty=0.1128(Affirmative) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```


The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 854

A DNA sequence (GBSx0906) was identified in *S.agalactiae* <SEQ ID 2581> which encodes the amino acid sequence <SEQ ID 2582>. This protein is predicted to be nad(p)h nitroreductase ydgi. Analysis of this protein sequence reveals the following:

```

10   Possible site: 38
    >>> Seems to have no N-terminal signal sequence
        INTEGRAL    Likelihood = -1.81    Transmembrane  127 - 143 ( 126 - 143)

    ----- Final Results -----
15         bacterial membrane --- Certainty=0.1723(Affirmative) < succ>
           bacterial outside --- Certainty=0.0000(Not Clear) < succ>
           bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

20   >GP:CAC09964 GB:AX033132 unnamed protein product [Bacillus subtilis]
        Identities = 62/204 (30%), Positives = 106/204 (51%), Gaps = 11/204 (5%)

    Query: 3   FLELNKKRHAVKHFNDDKPVDFKDVRTAI-EIA'ATLAPSANNIQPWKFVVVQ--EKKSALAE 59
               F+E+ K R ++++++      K+  T I E AT APS+ N QPW+F+V+   E K  LA
25   Sbjct: 7   FMEIMKGRRSIRNYDPAVKISKEEMTEILEEATTAPSSVNAQPWRFLVIDSPEGKEKLAP 66

    Query: 60   GLPESNCNQINQAQYVIALFTDTD---LGQRSRKIARIGRRSLPDDLIGYYMETLPPEY 115
               L  N  Q+  +  VIA+F D +   L +  K  +G  +P ++   +  L  +
30   Sbjct: 67   -LASFNQIQVTTSSAVIAVFADMMNADYLEEIIYKAVELG--YMPQEVKDRQIAALTAHF 123

    Query: 116  ALYSEKQTGEYLSLNAGIVAMNLVLALTDQGISSNMILGFDKAITNDVLEIDK-RFRPEI 174
               +   E + ++ G+V+M L+L      G  +N I G+DK   +   +DK R+ P +
35   Sbjct: 124  EKLPAQVNRETILIDGLVSMQLMLTARAHGYDTNPIGGYDKENIAETFGLDKERYVPVM 183

    Query: 175  LITVGYSDKVEPSYRLPVDHIIIE 198
               L+++G + ++   SYRLP+D I E
40   Sbjct: 184  LLSIGKAADEGYASYRLPIDTIAE 207

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2583> which encodes the amino acid sequence <SEQ ID 2584>. Analysis of this protein sequence reveals the following:

```

    Possible site: 38

    >>> Seems to have no N-terminal signal sequence
        INTEGRAL    Likelihood = -2.18    Transmembrane  127 - 143 ( 126 - 143)
45   ----- Final Results -----
           bacterial membrane --- Certainty=0.1871(Affirmative) < succ>
           bacterial outside --- Certainty=0.0000(Not Clear) < succ>
           bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
50

```

The protein has homology with the following sequences in the databases:

```

    >GP:CAC09964 GB:AX033132 unnamed protein product [Bacillus subtilis]
        Identities = 63/204 (30%), Positives = 109/204 (52%), Gaps = 11/204 (5%)

55   Query: 3   FLELNKKRHAIKTFNDQ-PVDYEDLRTAIEIATLAPSANNIQPWKFVVVQ--EKKAELAK 59
               F+E+ K R +I+ ++   +  E++   +E AT APS+ N QPW+F+V+   E K +LA

```

-936-

Sbjct: 7 FMEIMKGRRSIRNYDPAVKISKEEMTEILEEATTAPSSVNAQPWRFLVIDSPEGKEKLA- 65

Query: 60 GLPLA--NKVQVEQAQYVVALFSDTDLALRSRKIARIGVK--SLPDDLIGYYMETLPPRF 115
 PLA N+ QV + V+A+F+D + A +I V+ +P ++ + L F

5 Sbjct: 66 --PLASFNQTQVTSSAVIAVFADMMNADYLEEIIYSKAVELGYMPQEVKDRQIAALTAHF 123

Query: 116 AAFNEVQTGEYLAINAGIVAMNLVLSLTDQKIASNIIIGFDKSTTNEILDID-PRFRPEL 174
 E + I+ G+V+M L+L+ +N I G+DK E +D R+ P +

10 Sbjct: 124 EKLPAQVNRETILIDGGLVSMQLMLTARAHGYDTNPIGGYDKENIAETFGLDKERYVPVM 183

Query: 175 LITVGYSDEKPEPSYRLPVDEVIE 198
 L+++G + ++ SYRLP+D + E

Sbjct: 184 LLSIGKADEGYASYRLPIDTIAE 207

15 An alignment of the GAS and GBS proteins is shown below.

Identities = 157/200 (78%), Positives = 184/200 (91%)

Query: 1 MKFLELNKKRHAVKHFNDKPVDFKDVRTAIEIATLAPSANNIQPWKFVVVQEKKSALAEG 60
 MKFLELNKKRHA+K FND+PVD++D+RTAIEIATLAPSANNIQPWKFVVVQEK+ LA+G

20 Sbjct: 1 MKFLELNKKRHAIKTFNDQPVVDYEDLRTAIEIATLAPSANNIQPWKFVVVQEKKAELAKG 60

Query: 61 LPESNCNQINQAQYVIALFTDTDLGQSRKRIARIGRRSLPDDLIGYYMETLPPRYALYSE 120
 LP +N Q+ QAQYV+ALF+DTDL RSRKIARIG +SLPDDLIGYYMETLPPR+A ++E

25 Sbjct: 61 LPLANKVQVEQAQYVVALFSDTDLALRSRKIARIGVKSLPDDLIGYYMETLPPRFAAFNE 120

Query: 121 KQTGEYLSLNAGIVAMNLVLALTDOGISSNMILGFDKAITNDVLEIDKRFRPEILLITVGY 180
 QTGEYL++NAGIVAMNLVL+LTDQ I+SN+ILGFDK+ TN++L+ID RFRPE+LITVGY

Sbjct: 121 VQTGEYLAINAGIVAMNLVLSLTDQKIASNIIIGFDKSTTNEILDIDPRFRPEILLITVGY 180

30 Query: 181 SDEKVEPSYRLPVDHIIIEKR 200
 SDEK EPSYRLPVD +IE+R

Sbjct: 181 SDEKPEPSYRLPVDEVIERR 200

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 35 vaccines or diagnostics.

Example 855

A DNA sequence (GBSx0907) was identified in *S.galactiae* <SEQ ID 2585> which encodes the amino acid sequence <SEQ ID 2586>. Analysis of this protein sequence reveals the following:

Possible site: 37
 40 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2895(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 45 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC45369 GB:U78036 dipeptidase [Lactococcus lactis]
 Identities = 312/474 (65%), Positives = 370/474 (77%), Gaps = 11/474 (2%)

50 Query: 2 TIDFRAEVDKRDALMDDLINLLRINSERDDSQADAEHPFGPGPVKALEFFLEMAERDGY 61
 TIDF+AEV+KRDALM+DL +LLRI+S D ADAE+PFGGPGP KAL+ FL++AERDGY

Sbjct: 3 TIDFKAEEVKRDALMEDLFSLLRIDSAMDMEHADAENPFGPGPRKALDAFLKIAERDGY 62

55 Query: 62 ETKNVNDNYAGHFTFGQGE---EELGIFGHLDVVPAGSGWDITDYPYEPVIKDNRLYARGSS 117
 TKN DNY GHF + G E LGI GHLDVVPAGSGWD++P+EP I++ LYARG+S

Sbjct: 63 TTKNVDNYVGHFEYENGANADAENVLGIIGHLDVVPAGSGWDSNPFEPFIRNGNLYARGAS 122

60 Query: 118 DDKGPTMACYALKI+KELGLPTSKKVRVVGTDDESGWGDMDYYFEHVGLPKPDFGFSP 177
 DDKGPT+ACYALKI+KEL LP SKK+RF+VGT+EE+GW DMDYYFEH LP PDFGFSP

Sbjct: 123 DDKGPTVACYALKILKELNPLSKKIRFIVGTNEETGWADMYYFEHCELPLPDFGFSP 182

-937-

5 Query: 178 DAEFPIINGEKG NITEYLHFSGENKGAVRLHSFSGGLRENMPESATARFTSHLDQTTLG 237
DAEFPIINGEKG NITEYLHFSG+N G V LHSF GL ENMPESATA + D L
Sbjct: 183 DAEFPIINGEKG NITEYLHFSGKNAGQVVLHSFKAGLAENMPESATAVISGAKD---LE 239

10 Query: 238 ASLADFASKH---NLKAELSVEDEQYTATVYKSAHGSTPQEGVNGATYLLALVLSQDFDE 294
A+L F ++H NL+ +L D + T T+YKSAHG+ P++G+NGATYL L+L+QDFD
Sbjct: 240 AALEKFVAEHASKNLRFDLEEDGKATITLYGKSAHGAMPEKGINGATYLLTFLNQDFDA 299

15 Query: 355 QGTDAKTIQTKLEKLVGKVTLSDEHTPHYVPMDELVSTLLAVYEKQTGLKGHEQVI 414
QG + +Q L KL+GV +V LS H HTPHYVEM D LVSTL+ VYEK TGLKG+E +I
Sbjct: 359 QGNSPERMQEILAKLDGVVEVELSKHLHTPHYVMSDPLVSTLIDVYEKHTGLKGYETII 418

20 Query: 415 GGGTFGRLLERGVA GAMFPGDENTMHQANEYMPLENI FRSAATYAEAIYELIK 468
GGGTFGRLLERGVA GAMF G+ ++MHQANE P+ENI+++A IYAEAIYEL K
Sbjct: 419 GGGTFGRLLERGVA GAMFGE PDSMHQANEMKPVENIYKAAVIYAEAIYELAK 472

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2587> which encodes the amino acid sequence <SEQ ID 2588>. Analysis of this protein sequence reveals the following:

25 Possible site: 14
>>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.3107(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 361/467 (77%), Positives = 403/467 (85%)

35 Query: 2 TIDFRAEVDKRRKDALMDDLINLLRINSERDDSQADAHPFGPGPVKALEFFLEMAERDGY 61
TIDF+AEVDKRRK A++ DL++LLRINSERDD AD +HPFGPGPVKALE FL MAERDGY
Sbjct: 20 TIDFKA EVDKRRKAMLADLVLLRINSERDDQLADDKHPFGPGPVKALEHFLAMAERDGY 79

40 Query: 62 ETKNVDNYAGHFTFGQGEELGIFGHLDVVPAGSGWDTDPYEPVIKDNRLYARGSSDDKG 121
+T+N+DNYAG F FGQG+E LGTFGHLDVVPAGSGWDTDPYEPVIK+R+YARGSSDDKG
Sbjct: 80 KTRNIDNYAGDFEFGQGEDEVLGIFGHLDVVPAGSGWDTDPYEPVIKDDRIYARGSSDDKG 139

45 Query: 122 PTMACYYALKI IKELGLPTS KKVRFVVGTDDESGWGDMDYYFEHVGLPKPDFGFS PDAEF 181
PTMACYYALKI IKELGLP SKKVRF+VGTDESGWGDMDYYF H GL PDFGFS PDAEF
Sbjct: 140 PTMACYYALKI IKELGLPVSKKVRFIVGTDESGWGDMDYYFAHNGLKNPDFGFS PDAEF 199

50 Query: 182 PIINGEKG NITEYLHFSGENKGAVRLHSFSGGLRENMPESATARFTSHLDQTTLGASLA 241
PIINGEKG NITEYLHF+G+NKGA LH F GGLRENMPESATA T+ D L A+L
Sbjct: 200 PIINGEKG NITEYLHFAGDNKGAFVLRHQGLRENMPESATAVITAPHDLVDLEAALE 259

55 Query: 242 DFASKHNLKAELSVEDEQYTATVYKSAHGSTPQEGVNGATYLLALVLSQDFDEGEARAPL 301
F S+H +K + D + T+ GKSAHGSTP+ GVNGAT LA +L+QF FEG A+ +L
Sbjct: 260 QFLSEHGKVGSMKATDGKIEVTIIGKSAHGSTPEAGVNGATLLAKFLNQFTFEGA AKDYL 319

60 Query: 302 DVTANIIHEDFSGEKLG VAYEDDCMGPLSMNAGVFQFDETNDNTIALNFRYPQGTDAKT 361
V ++HEDF+ EKLGA+Y DD MG LSMNAGVF FD + DNTIALNFRYP+GTDA T
Sbjct: 320 HVAGEVLHEDFAAEKLG LAYTDDRMGALSMNAGVFTFDSQSADNTIALNFRYPKGTDAAT 379

65 Query: 362 IQTKLEKLVGKVTLSDEHTPHYVPMDELVSTLLAVYEKQTGLKGHEQVIGGGTFGR 421
++ LEKL G+ KV+LS+HEHTPHYVPMDELV+TLLAVYEKQTGLKG+EQVIGGGTFGR
Sbjct: 380 LKAGLEKLPGLTKVLSSEHEHTPHYVPMDELVATLLAVYEKQTGLKGYEQVIGGGTFGR 439

Query: 422 LLERGVA GAMFPGDENTMHQANEYMPLENI FRSAATYAEAIYELIK 468
LLERGVA+GAMFPGDENTMHQANEYMPLENI+RSAATYAEAIYELIK
Sbjct: 440 LLERGVA GAMFPGDENTMHQANEYMPLENIYRSAATYAEAIYELIK 486

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 856

- 5 A DNA sequence (GBSx0908) was identified in *Sagalactiae* <SEQ ID 2589> which encodes the amino acid sequence <SEQ ID 2590>. Analysis of this protein sequence reveals the following:

Possible site: 30
>>> Seems to have no N-terminal signal sequence

10 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.5598 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 15 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC21888 GB:U32707 H. influenzae predicted coding region
HI0220.2 [Haemophilus influenzae Rd]
Identities = 123/192 (64%), Positives = 160/192 (83%), Gaps = 1/192 (0%)

20 Query: 1 MTDLEKIIKAISDSQNQNYTENGIDPLFAAPKTARINIVGQAPGLKTQEARLYWKDKSG 60
+ +L++I +I +D QN+++TE GI PLF+APKTARINIVGQAPGLK +++RLYW DKSG
Sbjct: 21 LKNLDEITSSIIADPQNKDFTERGIFPLFSAPKTARINIVGQAPGLKAEQSRLYWNDKSG 80

25 Query: 61 DRLRQWLGVDEETFYHSGKFAVLPLDFYYPGKSGDLSPRKGFAEKWHPLILKEMPVQ 120
DRLR+WLGV D + FY+SG FAVLP+DFYYPG KSGDL PR+GFAB+WHP+IL +PN+Q
Sbjct: 81 DRLREWLGV D YDYFYNSGIFAVLPMDFYYPGKSGDLP RPQGFAERWHEMILGNL PNIQ 140

Query: 121 LTLVGQYTQKYYLGSSAHKNLTETVKAYKDYLPDYLPVHPSPRNQIWLKKNPWFEKDL 180
LT+L+GQY QKYYL + N+T TVK Y+ +LP ++PLVHPSPRNQ+W+ KNPWFE+ +
30 Sbjct: 141 LTILIGQYAQKYYLPEN-KDNVTNTVKNYRQFLPHFPLVHPSPRNQLWVTKNPWFEQV 199

Query: 181 IVDLQKIVADIL 192
I +LQ +V I+
Sbjct: 200 IPELQILVKQII 211

- 35 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2591> which encodes the amino acid sequence <SEQ ID 2592>. Analysis of this protein sequence reveals the following:

Possible site: 47
>>> Seems to have no N-terminal signal sequence

40 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.3740 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 45 An alignment of the GAS and GBS proteins is shown below.

Identities = 122/189 (64%), Positives = 150/189 (78%)

50 Query: 4 LEKIIKAISDSQNQNYTENGIDPLFAAPKTARINIVGQAPGLKTQEARLYWKDKSGDRL 63
++ + KAI +D N +YTE GI PL+ AP+TARI IVGQAPG+ Q +LYW D+SG RL
Sbjct: 1 MDDLTKAIMADEANLSYTERGIFPLYDAPQTARIIVGQAPGIVAGGTKLYWND RSGIRL 60

Query: 64 RQWLGVDEETFYHSGKFAVLPLDFYYPGKSGDLSPRKGFAEKWHPLILKEMPVQ LTL 123
R WLGV D +TFYHSG F ++P+DFYYPGKSGDL PR+GFA KWHP + MP V+LT+
55 Sbjct: 61 RDWLGV DNDTFYHSG LFGIIPMDFYYPGKSGDLP RPREGFAAKWHPLRLALMP EVELTI 120

Query: 124 LVGQYTQKYYLGSSAHKNLTETVKAYKDYLPDYLPVHPSPRNQIWLKKNPWFEKDLIVD 183
LVG+Y Q +YLG+ A+K LTFIV+ ++DYLPDY PLVHPSPRNQ+WL KNPWFE+DL+
Sbjct: 121 LVGRYAQDFYLGKAYKTLTETVRHFEDYLPDYLPVHPSPRNQLWLAKNPWFQD LLEI 180

-939-

Query: 184 LQKIVADIL 192
 LQK V IL
 Sbjct: 181 LQKRVEAIL 189

5

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 857

A DNA sequence (GBSx0909) was identified in *S.agalactiae* <SEQ ID 2593> which encodes the amino acid sequence <SEQ ID 2594>. Analysis of this protein sequence reveals the following:

Possible site: 16
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 15 bacterial cytoplasm --- Certainty=0.4178 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

20 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 858

A DNA sequence (GBSx0910) was identified in *S.agalactiae* <SEQ ID 2595> which encodes the amino acid sequence <SEQ ID 2596>. Analysis of this protein sequence reveals the following:

Possible site: 45
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 30 bacterial cytoplasm --- Certainty=0.2779 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9899> which encodes amino acid sequence <SEQ ID 9900> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAD35886 GB:AE001748 conserved hypothetical protein [Thermotoga maritima]
 Identities = 36/124 (29%), Positives = 58/124 (46%), Gaps = 3/124 (2%)

40 Query: 19 VPTKELLADYFNRMEFAIGRVEAHVLAHFDYGFRLNLDVEDLKPFFETQLKRIFIKMLSK 78
 +P EL DY R F + RV+ H LAH DY R D K +++I + ++
 Sbjct: 98 LPPDELARDYLERTLFVMERVKFHTLAHLDPARYAKAD---FKANRDLEIKILVFLVKN 154

45 Query: 79 GLAFELNTKSLYLYGNEKLYRYALEILKQLGCKQYSIGSDGHIPEHFCYEFDRLOGLLKD 138
 A E+NT L+ +G + +E+ LG + +IGSD H +H + + LK
 Sbjct: 155 EKALEINTAGLFKHKGNPDYWIVEMYYDLGGRVVTIGSDAHESQHIGRGIEEVMRELKK 214

Query: 139 YQID 142
 + +

50 Sbjct: 215 FNFE 218

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 859

A DNA sequence (GBSx0911) was identified in *S. agalactiae* <SEQ ID 2597> which encodes the amino acid sequence <SEQ ID 2598>. This protein is predicted to be alkaline amylopullulanase (pula). Analysis of this protein sequence reveals the following:

```

Possible site: 41
>>> Seems to have no N-terminal signal sequence
    INTEGRAL    Likelihood = -10.08    Transmembrane 1225 -1241 (1222 -1247)
10    INTEGRAL    Likelihood = -2.44    Transmembrane 19 - 35 ( 18 - 36)
    INTEGRAL    Likelihood = -0.11    Transmembrane 1146 -1162 (1146 -1162)

----- Final Results -----
15    bacterial membrane --- Certainty=0.5034(Affirmative) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAG33958 GB:AF217414 pullulanase [Streptococcus pneumoniae]
20    Identities = 641/1311 (48%), Positives = 854/1311 (64%), Gaps = 88/1311 (6%)

Query: 1    MKRKDLFGDKQTQYTIRKLSVGVA SVATGVCIFLHSPQVFAEEVSASPANTAIAESNINQ 60
          M++   +K+  Y+IR L  G  SV  G  +  L                      A+A  I+
Sbjct: 1    MRKTPSHTEKKMVYSIRSLKNGTGSVLIGASLVL-----LAMATPTISS 44

25    Query: 61    VDNQQSTNLKDDINSNSETVTPSPDMPDTKQLVSDETDTQKGVTEPDKATSLLEENKG-P 119
          ++  +TN  +  N N+ T+  P  + DT  +      +  ++ P  A  + LE+ +  P
Sbjct: 45    DESTPTTN--EPNNRNTTTLAQP--LTDT---AAGSGKNESDISSPGNANASLEKTEEKP 97

30    Query: 120   VSDKNTLDLKVAPSTLQNTPKDTSQAIGAPSPTLKVANQAPRIENGYFRLHLKELPQGHP 179
          ++  T    A    Q  D++S+  + SP              IE+ YFR+H+K+LP+ +
Sbjct: 98    ATEPITPAASPADPAPQTGQDRSSEPTTSTSPVTTETKAEPIEDNYFRIHVKKLPEENK 157

35    Query: 180   VESTGLWIWGDVDQPSNWPNGAI PMTDAKKDDYGYVDFKLSEKQKQISFLINN KAGT 239
          ++  GLW  W  DV++PS NWPNGA+  DAKKDDYGY+D KL  +Q  K+ISFLINN AG
Sbjct: 158    -DAQQLWTWDDVEKPSNWPNGALSFKDAKKDDYGYLDVKLKGEQAKKISFLINNTAG 216

40    Query: 240   NLSGDHHIPLLRPEMNQVWIDEKYGHTYQPLKEGYVRINYLSSSSNYDHL SAWLFKDVA 299
          NL+GD  +  L  P+MN+ W+D+ Y  +Y+P  G VR+NY  +  NYD  S W  +  DV
Sbjct: 217    NLTGDKSVEKLVPKMNEAWLQDQYKVFSEYEPQAGTVRVNYYRTDGN YDKKSLWYWG DVK 276

45    Query: 300   TPSTI-WPDGSNFVNQGLYGRYIDVSLKTNAKEIGFLILDESKTGDAVKVQPN DYVFRDL 358
          PS+  WPDG++F  G  YGRYID+ L  A+E  GFL+LDESK  GD  VK++  +Y  F  DL
Sbjct: 277    NPSSAQWPDGTDFTATGKYGRYIDIPLEAAREFGFLLLDESKQGDDVKIRKENYKFTDL 336

50    Query: 359   ANHNQIFVKDKDPKVYNNPYIIDQVQLKDAQQIDLTISIQASFTTLDGVDKTEILKELKVT 418
          NH+QIF+KD D  +Y  NPYY+  +++  AQ  +  +SI++SF+TL  G  K  +ILK  +T
Sbjct: 337    KNHSQIFLKDDDESIYTNPHYVHDIRMTGAQHVGTS SSISSFSSTLVGAKKEDILKHSNIT 396

55    Query: 419   DKNQNAIQISDITLDTSKSLIIKGFNPKQGHEFNISYNGNNVMTRQSWEFKDQLYAYSG 478
          +  N  +  I+D+  +D  +  +  GDF+  +  +  +SYN  +  T+  SW  KD+  Y+Y  G
Sbjct: 397    NHLGNKVITITDVAIDEAGKKVYISGDFSDTKHPYTVSYNSDQFTTKTSWRLKDETYSDG 456

60    Query: 479   NLGAVLNQDGSKVEASLWSPSADSVTMIIDKDNQNRVVATTPLMKNNKGWVQTILDT-- 536
          LGA  L  ++G  +V+  +LWSPSAD  V++++YDK++  ++VV  T  L  K  +G  W+  LD+
Sbjct: 457    KLGADLKEEGKQVDLTLWSPSADKSVVVYDKNDPDKVVGTVALEKGERGTWKQTL DSTN 516

Query: 537   KLGIKNYTGYYLYEIKRGKDKVKILDYAKSLAEWDSNT--VNDDIKTAKAAFVNPSQL 594
          KLGI  ++TGYYY  Y+I+R  V  LDPYAKSLA  W+S+  ++D  K  AKAAPV+P++L
Sbjct: 517   KLGITDFTGYYYQYQIERQGTVLALDPYAKSLA AWNSDDAKIDDAHKVAKA AFVDPAKL 576

Query: 595   GPQNLSFAKIANFKGRQDAVIYEAHVRDFTSDRSLDGKLNQFGTFAAFSEKLDYLQKLG 654

```

-941-

5
 10
 15
 20
 25
 30
 35
 40
 45

Sbjct: 577 GPQ+L++ KI NFK R+DAVIYEAHVDRFTSD ++ L FGTF AF EKLDYL+ LG
 GPQDLTYGKIHNFKTREDAVIYEAHVDRFTSDFAIAKDLTKPFGTFEAFIEKLDYKDLG 636

Query: 655 VTHIQLLPVLSYFYVNEMDKSRSTA-YTSSDNNYNWGYDPQSYFALSGMYSEKPKDPSAR 713
 VTHIQLLPVLSY++VNE+ + Y SS++NYNWGYDPQ+YF+L+GMYS PK+P R
 Sbjct: 637 VTHIQLLPVLSYFYVNEELKNHEHLSYASSNSNYNWGYDPQNYFSLTGMYS S DPKNFEKR 696

Query: 714 IAELKQLIHDHDKRGMGVILDVVYNHTAKTYLFEDIEPNYVHFMNEDGSPRESFGGGRLG 773
 IAE K LI++IHKRGMG ILDDVVYNHTAK +FED+EPNYYHFM+ DG+PR SFGGGRLG
 Sbjct: 697 IAEPKLINEIHKRGMGAILDDVVYNHTAKVDIFEDLEPNYVHFMADGTPRTSFGGGRLG 756

Query: 774 TTHAMSRRLVDSIKYLTSEFKVDGFRFDMMGDHDAAAIELAYKEAKAINPNMIMIGEGW 833
 TTH M++R+LVDSIKYL +KVDGFRFDMMGDHDAA+IE AYK A+A+NPN+IM+GEGW
 Sbjct: 757 TTHHMTKRLLVDSIKYLVDTYKVDGFRFDMMGDHDAAASIEEAYKAARALNPNLIMLGEGW 816

Query: 834 RTFQSDQGQPVKPADQDWMKSTDTVGVSDDIRNSLKSGFPNEGTPAFITGGPQSLQGIF 893
 RT+ GD+ P K ADQDMK TDTV VFSDDIRN+LKSG+PNEG PAFITGG + + IF
 Sbjct: 817 RTYAGDENMPTKAADQDWMKHTDTVAVFSDDIRNNLKSGYPNEGQPAFITGGKRDVNTIF 876

Query: 894 KNIKAQPGNFADSPGDVQYIAAHNLTLDHVIASINKDPKVAEE--EIHRLRLGNV 951
 KN+ AQP NFEADSPGDV+QYIAAHNLTLD+IA+SI KDP AE EIHRLRLGN+
 Sbjct: 877 KNLIAQPTNFEADSPGDVQYIAAHNLTLDIIAQSIKKDPSKAENYAEIHRRLRLGNL 936

Query: 952 MILTSQGTAFIHSGQEYGRKRLNPDYMTKVSDDKLPNKATLIEAVK----EYPYFIHD 1007
 M+LT+QGT FIHSGQEYGRK+ NP Y T V++DK+PNK+ L+ +YYPYFIHD
 Sbjct: 937 MVLTAQGTFFIHSGQEYGRKQFRNPAYRTPVABDKVPNKSHLLRDKGNPFDPYFYFIHD 996

Query: 1008 SYDSSDAINHFDWAAATDNNKHPISTKTQAYTAGLITLRRSTDAFRKLSKAEIDREVSLI 1067
 SYDSSDA+N FDW ATD +P + K++ Y GLI LR+STDAFR S +I V LI
 Sbjct: 997 SYDSSDAVNKFDWTKATDGKAYPENVKSRDYMKGILALRQSTDAFRKLSLQDIKDRVHLI 1056

Query: 1068 TEVGQGDIEKDLVIAIYQITDSKGDYAVFVNADSKARNVLLGEKYKHLKQVIVDADQ 1127
 T GQ ++++D+VI YQ GDIYAVFVNAD KAR LG + HL +V+ D +Q
 Sbjct: 1057 TVPGQNGVEKEDVVIGYQITAPNGDIYAVFVNADEKAREFNLTAFALRNAEVLADENQ 1116

Query: 1128 AGIKPISTPRGVHFEKDSLLIDPLTAIVIKVGKVAPS-----PKEELQAD 1172
 AG I+ P+G+ + + L ++ LTA V++V + S P+ + +A
 Sbjct: 1117 AGSVGIANPKGLEWTEKGLKLNALTATVLRVSONGTSHESTAEKPDSTPSKPEHQNEAS 1176

Query: 1173 YPKTQ-----SFKESKTVEKVNRIANKT-----SITPVVSKKADS 1207
 +P Q + ++K + N+ + T S+ V K++
 Sbjct: 1177 HPAHQDPAPPEARPDSTKPDADAKVADAENKPSQATADSQAEPQAEQAASSVKEAVRKESVE 1236

Query: 1208 YLTNE-----ANLPKTGDKSSKILSVVGISILASLLALVGLSLKRN 1249
 + E A LP TG K+ L GIS+LA LL L G LK +
 Sbjct: 1237 NSSKENISATPDRQAEPLNTGIKNENKLLFAGISLLA-LLGL-GFLLKNKK 1285

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2599> which encodes the amino acid sequence <SEQ ID 2600>. Analysis of this protein sequence reveals the following:

50 Possible site: 35
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -10.83 Transmembrane 1153 -1169 (1148 -1171)
 INTEGRAL Likelihood = -1.97 Transmembrane 29 - 45 (28 - 46)

55 ----- Final Results -----
 bacterial membrane --- Certainty=0.5331(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

60 A related sequence was also identified in GAS <SEQ ID 9125> which encodes the amino acid sequence <SEQ ID 9126>. Analysis of this protein sequence reveals the following:

Possible cleavage site: 26
 >>> Seems to have an uncleavable N-term signal seq

-942-

----- Final Results -----

bacterial membrane --- Certainty= 0.533 (Affirmative) < succ>
 bacterial outside --- Certainty= 0.000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty= 0.000 (Not Clear) < succ>

5

LPXTG motif: 1133-1137

An alignment of the GAS and GBS proteins is shown below.

Identities = 715/1097 (65%), Positives = 872/1097 (79%), Gaps = 21/1097 (1%)

10 Query: 156 ANQAPRIENGYFRLHLKELPQGHEVESTGLWINGDQDPSSNWPNGAIPMTDAKKDDYGY 215
 AN A E+ + R+H K LP G + S GLW+WGDVDQPS +WPNGAI MT AKKDDYGY
 Sbjct: 95 ANPASTAEH-HLRMHFKTLPAGESELGSLGLVWVGVDQPSKDWPNGATMTAKKDDYGY 153

15 Query: 216 YVDFKLSEKQKQKQISFLINNKAGTNLSGDHHPILRPEMNQVWIDEKYGHTHYQPLKEGY 275
 Y+D L+ K R+Q+S+LINNKAG NLS D HI LL P+MN+VWIDE Y H Y+PLK+GY
 Sbjct: 154 YLDVPLAAKHRQVSYLINNKAGENLSKQDHISLLTPKMNEVWIDENYHAHAYRPLKKG Y 213

20 Query: 276 VRINYLSSSSNYDHLASAWLFKDVATPSTTWPDGSNPFVNGGLYGRYIDVSLKTNKEIGFL 335
 +RINY + S +YD+L+ W FKVDP TP+T WP+G + ++G YG Y+DV LK A EIGFL
 Sbjct: 214 LRINYHNQSGHYDNLAVWTFKDVKTPTTDWPNGLDLSHGKHYGAYVDVPLKEGANEIGFL 273

25 Query: 336 ILDESKTGDAVKVQPNQYVFRDLANHNQIFVKDKDPKVYNNPYYIDQVQLKDAQQIDLTS 395
 ILD+SKTGDA+KVQP DY+F++L NH Q+FVKD DPKVYNNPYYIDQV LK A+Q
 Sbjct: 274 ILDKSKTGDAIKVQPKDYLFKELDNHTQVFKDTPDKVYNNPYYIDQVSLKGAEQITPNE 333

30 Query: 396 IQASFTTLDGVDKTEILKELKVTDKNQNAIQISDITLDTSKSLIIKGFNPKQGHFNIS 455
 I+A FTTLDG+D+ + + +K+TDK + I ++TLD KS++ +KGDF + + ++
 Sbjct: 334 IKAIFTTLDGLDEDVAVKQNKITDKAGKTVAIDELTLDLDRDKSVMTLKGDFKAQGAVYTVT 393

35 Query: 456 YNGNNVMTRQSWEFKDQLYAYSGNLGAVLNQDGSKEASLWSPSADSVTMIYDKDNQNR 515
 + + + RQSW+ KD+LYAY G LGA L +DGS V+ +LWSPSAD+V +++YDK +Q R
 Sbjct: 394 FGEVSQVARQSWQLKDKLYAYDGEGLGATLAKDGS-VDLALWSPSADTVKVVVYDKQDQTR 452

40 Query: 516 VVATTPLMKNNKGVWQTIL--DTKLGIKNYTGYYLYEIKRGDKVKILDPYAKSLAEWD 573
 VV L K++KGVW+ L D+ GI +YTGYYLYEIRG++KV +LDPYAKSLA W+
 Sbjct: 453 VVGQADLTKSDKGVWRAHLTSDSVKGISDYTGYYLYEITRGQEKVMVLDPYAKSLAAWN 512

45 Query: 574 SNTVNDDIKTAKAAFVNPSQLGPQNLSFAKIANFKGRQDAVIYEAHVDRFTSDRSLDGKL 633
 T DDIKTAKAAF++PS+LGP L FAKI NFK R+DA+IYEAHVDRFTSD++L+GKL
 Sbjct: 513 DATATDDIKTAKAAFIDPSKLGPTGLDFAKINNFKKREDAIYEAHVDRFTSDKALEGKL 572

50 Query: 634 KNQFGTFAAFSEKLDYLQKLGVTHIQLLPVLVSIFYVNMEDKSRSTAYTSSDNNYNWGYDP 693
 + FGTF+AF E+LDYL+ LGVTH+QLLPVLVSIFY NE+DKSRSTAYTSSDNNYNWGYDP
 Sbjct: 573 THPFGTFSAFVEQLDYLKDLGVTHVQLLPVLVSIFYANELDKSRTAYTSSDNNYNWGYDP 632

55 Query: 694 QSYFALSGMYSEKPKDPSARIAELKQLIHDIHKRGMGVILDVVYNHTAKTYLFEDIEPNY 753
 Q YFALSGMYS P DP+ RIAELK L+++IHKRGMGVI DVVYNHTA+TYLFED+EPNY
 Sbjct: 633 QHYFALSGMYSANPNDFALRIABLKNLVNEIHKRGMGVIPDVVYNHTARTYLFEDLEPNY 692

60 Query: 754 YHFMNEDGSPRESFGGRLGTTHAMSRRLVDSIKYLTSEFKVDGFRFDMMGDHDAAAIE 813
 YHFMN DG+ RESFGGRLGTTHAMSRRLVDSI YLT EFKVDGFRFDMMGDHDAAAIE
 Sbjct: 693 YHFMNADGTARESFGGRLGTTHAMSRRLVDSITTYLTREFKVDGFRFDMMGDHDAAAIE 752

65 Query: 814 LAYKEAKAINPNMIMIGEGWRTFQSDQGPVKPADQDWMKSTDTVGVSDDIRNLSKSGF 873
 A-K AKAINPN IMIGEGWRT+QGD+G+ ADQDWMK+T+TVGVFSDDIRN+LKSGF
 Sbjct: 753 QAFKAAKAINPNTIMIGEGWRTYQDEGKKEIAADQDWMKATNTVGVSDDIRNLTLSKSGF 812

Query: 874 PNEGTPAFITGGPQSLQGIFKNIAQPGNFEADSPGDVVQYIAAHDNLTLDHVIKASINK 933
 PNEGTPAFITGG ++L+G+FK IKAQPGNFEAD+PGDVVQYIAAHDNLTLDHVIKASINK
 Sbjct: 813 PNEGTAAPITGGAKNLEGLFKTIKAQPGNFEADAPGDVVQYIAAHDNLTLDHVIKASINK 872

Query: 934 DPKVAEEBIIHRRRLRLGNVMILTSQGTAFIHSQGEYGRTRKLLNPDYMTKVSDDKLPNKAT 993
 DPKVAEEBIIH+R+RLGN MILT+QGTAFIHSQGEYGRTRK+LLNPDY TK SDDK+PNKAT
 Sbjct: 873 DPKVAEEBIIHRRIRLGNMILTALQGTAFIHSQGEYGRTRKLLNPDYKTKASDDKVPNKAT 932

Query: 994 LIEAVKEYPYFIHDSYDSSDAINHFDWAAATDNNKHPISTKTQAYTAGLITLRRSTDAFR 1053

10

15

20 A related GBS gene <SEQ ID 8673> and protein <SEQ ID 8674> were also identified. Analysis of this protein sequence reveals the following:

25

30

35

40

The protein has homology with the following sequences in the databases:

50

55

60

65

-944-

```

      70      80      90
1746      1776      1806      1836
-----YAYSGNIGAVLNQDGSKEASLWSPSADSVTMIYDKDN
5      | | | | | : : | : | | : | : :
FIEIKDGKAEVWILQGV~~~~ELIEGYKPARVIMMELDDYYDGELGAVYSPE--KTIFRVWSPVSKWVKVLLFKNGE
      110      210      220      230      240      250
1866      1896      1926      1956      1986      2016      2046      2076
10      QNRVVATPLMKNNKGVWQITILDTKLGIKNYTGYYLYEIKRGKDKVKILDPAKSLAEWDSNTVNDIKTAKAAFVNPS
      : : : | | : : : | : | : : : : : | : | : : : : :
DTEPYQVNVMEYKGNVWEAVVEGDL-----DGVFYLQLENYKIRTTVDPYSKAVYA-----NSKKSAVVNLA
      270      280      290      300      310      320
2106      2136      2166      2196      2226      2253      2283
15      QLGFPQNLFAKIANFKGRQDAVIYEAHVRDFTSDRSLDGKLNQFGTFAAFSEK-----LDYLQKLGVTTHIQLL
      : | : : : | : | : | : | : : : : : : : : : : : :
RTNPEGWENDRGPKIEGYEDAIYEHADITG--LENSGVKNK-GLYLGLTENTKPGPGGVTTGLSHLVELGVTHVHIL
      330      340      350      360      370      380      390
2313      2343      2373      2403      2433      2463      2493
20      PVLSEYFVNEMDKSRSTAYTSSDNNYNWGYDPQSYFALSGMYSEKPKDPSARIAELKQLIHDHKGGMGVILDVVYNHT-
      : : : : | : | : | : | : | : | : | : : : : : :
PFFDFYTGDELDK-----DFEKYYNWGYDPLYFMVPEGRYSTDPKNPHTRIREVKEMVKALHKGIGVIMDMVPFPTY
25      410      420      430      440      450      460      470
2544      2574      2601      2631      2661      2691      2721      2751
30      --AKTYLFEDIEPNYHFMNEDGSP-RESFGGRLGTTHAMSRRLVDSIKYLTSEFKVDGFRFDMMGDHDAAAIELAYK
      : | : | : : : : | : | : : : : | : : : : : | : : : : :
GIGELSAFDQTPYYPYRIDKTGAYLNEGCGNVIASERPMRKPIVDVTYVWVKEYHIDGFRFDQMLIDKKTMLEVER
      480      490      500      510      520      530      540      550
2781      2811      2841      2871      2901      2931      2979
35      EAKAINPNMIMIGEGWRTFGDQGGQPVKPADQDMKSTDTVGVSDDIRNSLKS GFNEGTPAFITGG----PQSLQGIF
      | : | : : | | : : : | : : | : : : : | : | : : : :
ALHKIDPTIILYGEPW---GGWGAPIRFGKSD--VAGTHVAAFNDEFRDALRGSVFNPSVKGFVMGGYGKETKIKRGVV
      560      570      580      590      600      610      620
3030      3060      3084      3114      3144      3174      3204
40      KNIAQPG---NFEADSPGDVVQYIAAHDNLTLHD--VIAKSINKDPKVAEEIHRRLRLGNVMILTSQGTAFIHSGQEY
      : | : : : | : : : | : : : | : : : | : : : | : : :
GSINYDGKLIKSFALD-PEETINYAACHDNHTLWDKNYLAADKKKEWTEELKNAQKLAGAILLTSQGVPLHGGQDF
      640      650      660      670      680      690      700
3234      3264      3294      3324      3354      3384      3414      3444
45      GRTKRLNPNPYMTKVSDDKLPNKATLIEAVKEYPYFIHDSYDSSDAINHFDWAAATDNNKHPISTKTQAYTAGLITIRRS
      | | | : | : : : | : : : | : : : : : : : : : | | | :
CRTKN-----FNDNSYNAPISINGFDY-----ERKLQFIDVFNYHKGLIKLRKE
      710      720      730      740
3474      3504      3534      3564      3594      3624      3654
50      TDAFRKLSKAEI-----DREVSLITEVGQDIKEKDLVIAYQTIDSKGDIYAVFVNADSKARNVLLGEKYKHLK
      | | | : | : : : | : : : : | : : : : | : : : : | : : :
HPAFRLKNAEEIKKHLEFLPGRRIVAFMLKDHAGGDPWKDIVVIYN-----GNLEKTTYK-LPE
      760      770      780      790      800
3678      3708      3738      3768      3798      3828      3858      3888
55      GQ--VIVDADQAGIKPISTPRGVHFEKDLLIDPLTAIVIKVGKVAPSPKEELQADYPKTQSFKEKSTVEKVNRIANKTS
      | : | : : : | : : : : | : : : : | : : : : | : : :
GKWNVVVNSQKAGTEVIETVEG-----TIEDPLSAYVLYRE
      820      830      840
60

```

SEQ ID 2598 (GBS5) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 3 (lane 7; MW 134kDa).

65 The His-fusion protein was purified as shown in Figure 190, lane 7.

-945-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 860

A DNA sequence (GBSx0912) was identified in *S.agalactiae* <SEQ ID 2601> which encodes the amino acid sequence <SEQ ID 2602>. Analysis of this protein sequence reveals the following:

Possible site: 26

>>> Seems to have no N-terminal signal sequence

10	INTEGRAL	Likelihood = -10.72	Transmembrane	231 - 247 (228 - 251)
	INTEGRAL	Likelihood = -8.39	Transmembrane	50 - 66 (44 - 68)
	INTEGRAL	Likelihood = -6.74	Transmembrane	23 - 39 (20 - 41)
	INTEGRAL	Likelihood = -5.84	Transmembrane	173 - 189 (168 - 196)
	INTEGRAL	Likelihood = -4.41	Transmembrane	299 - 315 (297 - 318)
	INTEGRAL	Likelihood = -4.14	Transmembrane	115 - 131 (114 - 133)
15	INTEGRAL	Likelihood = -3.35	Transmembrane	80 - 96 (79 - 97)
	INTEGRAL	Likelihood = -0.48	Transmembrane	97 - 113 (97 - 113)

----- Final Results -----

	bacterial membrane	---	Certainty=0.5288(Affirmative)	< succ>
20	bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
	bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

A related GBS nucleic acid sequence <SEQ ID 8675> which encodes amino acid sequence <SEQ ID 8676> was also identified. Analysis of this protein sequence reveals the following:

SRCFLG: 0

25 McG: Length of UR: 19

Peak Value of UR: 3.08

Net Charge of CR: 1

McG: Discrim Score: 9.76

30 GvH: Signal Score (-7.5): -4.57

Possible site: 22

>>> Seems to have an uncleavable N-term signal seq

Amino Acid Composition: calculated from 1

	ALOM program	count: 7	value: -10.72	threshold: 0.0
35	INTEGRAL	Likelihood = -10.72	Transmembrane	217 - 233 (214 - 237)
	INTEGRAL	Likelihood = -8.39	Transmembrane	36 - 52 (30 - 54)
	INTEGRAL	Likelihood = -6.74	Transmembrane	9 - 25 (6 - 27)
	INTEGRAL	Likelihood = -5.84	Transmembrane	159 - 175 (154 - 182)
	INTEGRAL	Likelihood = -4.14	Transmembrane	101 - 117 (100 - 119)
40	INTEGRAL	Likelihood = -3.35	Transmembrane	66 - 82 (65 - 83)
	INTEGRAL	Likelihood = -0.48	Transmembrane	83 - 99 (83 - 99)
	PERIPHERAL	Likelihood = 0.26		136
	modified ALOM score:	2.64		
	icml HYPID: 7	CFP: 0.529		

45 *** Reasoning Step: 3

----- Final Results -----

	bacterial membrane	---	Certainty=0.5288(Affirmative)	< succ>
50	bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
	bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB08178 GB:AB036768 exfoliative toxin A [Staphylococcus hyicus]

55 Identities = 134/298 (44%), Positives = 197/298 (65%)

Query: 22 PLVMAGLVGLLALGNLLLEGYGTYYVRYCLGLVALVFVIFLIKGLKNKKESRKELSNPLI 81

PLV +GLVGLL LGNLL+ + G++A++ W+ L+ + N + +L++PL+

Sbjct: 7 PLVSSGLVLGLLGLGNLLKDVSLSLNALCGILAILVWLHLLYSMFNNVNVHVKVQLNSPLV 66

60 Query: 82 ASVFTTFFMAGMILSTYILLFRSLGIWVAVLSKGVWVLSFIALIIHMAIFSWKYLRFHFSM 141

-946-

+SVFTTFFM+G + +TY+ F S ++ L +W L I ++ HM IFS KYL+ FS+
 Sbjct: 67 SSVFTTFFMSGFLGTTYNLNTFFSHISFIHHLITPLWLLCLIGILTHMIIFSHKYLKSFSL 126
 Query: 142 ANLFPSWSVLYVGIGVASLTAPISGQFTIGKIVFWYGFIA TLVLLPFLFIKAYKIGLPSA 201
 N++PSW+VLY+GI +A LTAP+SG F IGK+ YGF+AT ++LP +F + L ++
 Sbjct: 127 ENVYPSWTVLYIGIATAGLTAPVSGYFFIGKLTVIYGFVATCIVLFLVFKRLKTYPLQTS 186
 Query: 202 VKPNITTICAPMSLITAGYVNSFVSPNRGLLLLLLIVMAQFLYFFILFQVPKLLIGDFTPG 261
 +KPN +TICAP SL+ A YV +F + +++L +++Q YF+I+IQ+PKLL F+P
 Sbjct: 187 IKPNTSTICAPFSLVAAAYVLAFFPAHDFVVILFLILSQVFYFIVFQLPKLLREPFSPV 246
 Query: 262 FSAFTFPLVISATSLKLSIQHLSLPVDIQGLVHFEIGTTTLIVMIVMVRVYIFFLRRTI 319
 FSAFTFPLVISAT+LK S+ L P GL+ FE T+IV V YI + +
 Sbjct: 247 FSAFTFPLVISATALKNSMPILIPPEIWNGLLMFETVLATVIVFRVFFGYIHLFLKPV 304

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2603> which encodes the amino acid sequence <SEQ ID 2604>. Analysis of this protein sequence reveals the following:

Possible site: 40
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -9.82 Transmembrane 169 - 185 (163 - 189)
 INTEGRAL Likelihood = -8.49 Transmembrane 50 - 66 (38 - 69)
 INTEGRAL Likelihood = -7.86 Transmembrane 228 - 244 (224 - 247)
 INTEGRAL Likelihood = -5.15 Transmembrane 288 - 304 (284 - 306)
 INTEGRAL Likelihood = -3.29 Transmembrane 108 - 124 (107 - 126)
 INTEGRAL Likelihood = -3.29 Transmembrane 140 - 156 (140 - 161)
 INTEGRAL Likelihood = -1.33 Transmembrane 84 - 100 (84 - 100)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.4927(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 138/305 (45%), Positives = 200/305 (65%), Gaps = 5/305 (1%)
 Query: 12 RYMKNWKEPPLVMAGLVGLLALGNLLEGYGTYVRYCLGLVALVFWIFLIKILKNKKE 71
 R +MK+ + PPLVM+GL IG L+ GNLL Y + Y L AL + L+ G+++N +
 Sbjct: 12 RTLMKHLKTPPLVMSGLALGTLISFGNLLATYVVSIFNYLGILAAALFIYGILLVGMVRNLND 71
 Query: 72 SRKELSNPLIASVFTTFFMAGMILSTYILLFRSLGIWVAVLSKGVWVLSFIALIIHMAIF 131
 ++ +L PLIASVF TFFM GM+LS+ L G W+ L+ WWL F+ ++ +A +
 Sbjct: 72 TKMQLRQPLIASVFPTFFMTGMLLSSLFLKVTG-GCWLGF LT---WWLFFLGNLVLIAYY 127
 Query: 132 SWKYLRHFSMANLFPSWSVLYVGIGVASLTAPISGQFTIGKIVFWYGFIA TLVLLPFLFI 191
 ++++ FS N+FPSWSVL+VGI +A+LTAP S QF +G+++FW + T V+LPF+
 Sbjct: 128 QYRFVFSFSDNVFPSWSVLVFGIAMAALAPASRQFLGQVIFWVCLLLTAVILPFFMAK 187
 Query: 192 KAYKIGLPSAVKPNITTICAPMSLITAGYVNSFVSPNRGLLLLLLIVMAQFLYFFILFQVP 251
 K Y IGL AV PNI+T CAP+SL++A Y+ +F P G+++ L+V +Q LY F++ Q+P
 Sbjct: 188 KTYGIGLGQAVMPNISTFCAPLSLLSASYLATFPRPQVGMVIFLLVSSQLLYAFVVVQLP 247
 Query: 252 KLLIGDFTPGFSAFTFPLVISATSLKLSIQHLSLP-VDIQGLVHFEIGTTTLIVMIVMVR 310
 +LL F PGFSAFTFP VISATSLK+++ L + Q L+ E+ T +V V
 Sbjct: 248 RLLNRFNPGFSAFTFFPVISATSLKMTLSFLGWQGLGWQVLLGEVLLATALVTYVYGA 307
 Query: 311 YIFFL 315
 Y+ FL
 Sbjct: 308 YLRFL 312

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 861

A DNA sequence (GBSx0913) was identified in *S.agalactiae* <SEQ ID 2605> which encodes the amino acid sequence <SEQ ID 2606>. Analysis of this protein sequence reveals the following:

```

5   Possible site: 28
    >>> Seems to have a cleavable N-term signal seq.

    ----- Final Results -----
          bacterial outside --- Certainty=0.3000(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
10         bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2607> which encodes the amino acid sequence <SEQ ID 2608>. Analysis of this protein sequence reveals the following:

```

15  Possible site: 28
    >>> Seems to have a cleavable N-term signal seq.

    ----- Final Results -----
          bacterial outside --- Certainty=0.3000(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
20         bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

25  Identities = 45/57 (78%), Positives = 53/57 (92%)

    Query: 1  MVKKFAFAKGIATGVVATAATLAGAFAIKKTIIEPEEEKIAFIEENRKKAAARKVS 57
              MVKK+ F KG+ATGV+ATAAT+AGA FA+KKTII+PEEEK AFIEENRKKAAAR+RV+
    Sbjct: 1  MVKKYQFVKGLATGVLATAATVAGAVFAVKKTIIDPEEEKAAFIEENRKKAAARRVA 57

```

30 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 862

A DNA sequence (GBSx0914) was identified in *S.agalactiae* <SEQ ID 2609> which encodes the amino acid sequence <SEQ ID 2610>. This protein is predicted to be tRNA isopentenylpyrophosphate transferase (miaA). Analysis of this protein sequence reveals the following:

```

35  Possible site: 20
    >>> Seems to have an uncleavable N-term signal seq

    ----- Final Results -----
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
40         bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

45 A related GBS nucleic acid sequence <SEQ ID 9897> which encodes amino acid sequence <SEQ ID 9898> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

    >GP:BA06085 GB:AP001515 tRNA isopentenylpyrophosphate transferase
      [Bacillus halodurans]
    Identities = 139/311 (44%), Positives = 200/311 (63%), Gaps = 21/311 (6%)
50  Query: 7  KIKLIAVVGPTAVGKTALGLIELAKTFNGEIIISGDSQQVYQKLDIGTAKASKEEQEAYHH 66
              K KL+A+VGPTAVGKT + LAK NGE+ISGDS QVY+ +DIGTAK + EE + HH
    Sbjct: 2  KEKLVAIVGPTAVGKTKTSVMLAKRLNGEIVISGDSMQVYRGMDIGTAKITAEMDGVPHH 61

```

-948-

Query: 67 LIDVREVNENYSVYDFVKEAKVAIDTIISKGIPIIVGGTGLYLQSLFEGYHLGGEVNQE 126
 LID+++ +E++SV DF A I I +G++P +VGGTGLY+ ++ ++LG E
 Sbjct: 62 LIDIKDPSESFSVADFQDLATPLITEIHERGRPLFLVGGTGLYVNAVHOFNLGDIRADE 121

5 Query: 127 TLMAYREKLE----SLSDDELFEKLT----EQSIIPQVNRRAIRALELAKF----- 171
 YR +LE S + L +KL+ + + I N RR IRALE+ K
 Sbjct: 122 D---YRHELEAFVNSYGVQALHDKLSKIDPKAAAAIHPNNYRRVIRALEIIKLTGKTUTE 178

10 Query: 172 -GNDLQNSESPYDVLLIGLNDDRQVLYDRINRRVLDLMDNGLLDEAKWLYD-NYPSVQAS 229
 + + SPY++++IGL +R VLYDRINRRVD M++ GL+DEAK LYD Q+
 Sbjct: 179 QARHEETPSPYNLVMIGLTMERDVLYDRINRRVDQMVEEGLIDEAKKLYDRGIRDQSV 238

15 Query: 230 KGIGYKELFPYFSKQIPLEEAVDKLKQNTRRFAKRLTWFRNRMNVEFIMVGEENYQQKI 289
 + IGYKE++ Y + LEEA+D LK+N+RR+AKRQLTWFRN+ NV + + + ++ +KI
 Sbjct: 239 QAIGYKEMYDYLGNVTLEEAIDTLKRNRRYAKRQLTWFRNKANVTWFDMDTVDVDFDKKI 298

Query: 290 KRKVSDFLSSK 300
 ++ +F++ K
 20 Sbjct: 299 -MEIHNFIAGK 308

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2611> which encodes the amino acid sequence <SEQ ID 2612>. Analysis of this protein sequence reveals the following:

Possible site: 27
 >>> Seems to have a cleavable N-term signal seq.

25 ----- Final Results -----
 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 30 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 202/296 (68%), Positives = 250/296 (84%)

35 Query: 5 MRKIKLIAVVGPTAVGKTALGI ELAKTFNGEIIISGDSQQVYQKLDIGTAKASKEEQEAY 64
 M KIK++ +VGPTAVGKTALGI LAK FNGEIIISGDSQQVY++LDIGTAKA++EEQE A
 Sbjct: 1 MTKIKIVVIVGPTAVGKTALGISLAKAFNGEIIISGDSQQVYRQLDIGTAKATQEEQRAAV 60

40 Query: 65 HHLIDVREVNENYSVYDFVKEAKVAIDTIISKGIPIIVGGTGLYLQSLFEGYHLGGEVN 124
 HHLID+REV E+YS YDFV++A+ +I I+S+GK+PIIVGGTGLYLQSL EGYHLGG+V+
 Sbjct: 61 HHLIDIREVTESYSAYDFVQDAQKSISDIVSRGKLPPIVGGTGLYLQSLLEGYHLGGQVD 120

45 Query: 125 QETLMAYREKLESLSDEDLFEKLT EQSIIPQVNRRAIRALELAKFGNDLQNSESPYDV 184
 QE + AYR +LE L D DL+E+L +I I QVNRRAIRALELA+F ++L+N+E+ Y+
 Sbjct: 121 QEAVKAYRNELEQLDDHDLIERLQVNNITIBQVNRRAIRALELAQFADELENAETAYEP 180

50 Query: 185 LLIGLNDDRQVLYDRINRRVLDLMDNGLLDEAKWLYDNYPSVQASKGIGYKELFPYFSKQ 244
 L+IGLNDDRQV+YDRIN+RV+ M++NGLL+EAKWLY++YP+VQAS+GIGYKELFPYF +
 Sbjct: 181 LIIGLNDDRQVIYDRINQRVNRMIENGLLEAKWLYEHYPTVQASRGIGYKELFPYFVGE 240

Query: 245 IPLEEAVDKLKQNTRRFAKRLTWFRNRMNVEFIMVGEENYQQKIKRKVSDFLSSK 300
 + L EA D+LKQNTRRFAKRLTWFRNRM V F + +Y Q + +V DFL K
 Sbjct: 241 MTLAEASDQLKQNTRRFAKRLTWFRNRMVSTAITAPDYPQVHDRVDRDFLGQK 296

55 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 863

A DNA sequence (GBSx0915) was identified in *S.agalactiae* <SEQ ID 2613> which encodes the amino acid sequence <SEQ ID 2614>. This protein is predicted to be hflX (hflX). Analysis of this protein
 60 sequence reveals the following:

-949-

Possible site: 35

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----

5 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

10 >GP:BAB06081 GB:AP001515 unknown conserved protein [Bacillus halodurans]
 Identities = 182/406 (44%), Positives = 254/406 (61%), Gaps = 12/406 (2%)

Query: 9 ERVILVGVELQDT--ENFEMSMEELASLAKTAGANVNVHYYQKRDKYDSKSFISGKLEE 66
 ERV LV +L + E FE S+EEL +L TA V++ QKR+ + ++IG GKL+E

15 Sbjct: 10 ERVFLVACQLPNMTDEQFEASLEELALTLTAQGTVIDRLTQKREAIEPATYIGRGKLDE 69

Query: 67 IKAIVEADEIDTVVNNRLTPRQNSNLEAELGVKVIDRMQLILDIFAMRARSHEGKLQVH 126
 + +E E D V+VN L+ Q NL LGV+VIDR QLILDIFA RA+S EGKLQV

20 Sbjct: 70 LAIKMEEQEADLVIVNGELSGSQVRNLTNRLGVRVIDRTQLILDIFAGRAKSREGKLQVE 129

Query: 127 LAQLKYMLPRLVGQGIMLSRQAGGIGSRGPESQLELNRRSIRHQISDIERQLKIVEKNR 186
 LAQL Y+LPR+VGQG LSR GGIG+RGPGGE++LE +RR IR +++DI++QLK K+R

25 Sbjct: 130 LAQLNYLLPRIVGQGQLSRLGGIGTRGPGETKLETDRRHIRKRMADIDKQLKHTVKHR 189

Query: 187 ETVRERRVDSTTFKIGLIGYTNAGKSTIMNVLTDDKQYEANELFATLDATTKQIYLQNF 246
 + R RR + TF+I L+GYTNAGKST++N LT YE + LFATLD T+++ L +

30 Sbjct: 190 DRYRARRERNQTFRIALVGYTNAGKSTLLNRLTASDSYEEDLLFATLDPMTRKMRPLSGM 249

Query: 247 QVTLTDTVGFIQDLPTLVAFAFKSTLEESRHVDLLFHVIDASDPNHEEHEKVMEILKDL 306
 +V L+DTVGFI LPT LVAAF+STLEE +H DLL HV+D S + H + V E+L L

35 Sbjct: 250 EVILSDTVGFINQLPTTLVAFAFRSTLEEVKHADLLLVVDRSSEQLOAHMETVSELLHQL 309

Query: 307 DMIDIPRLAIYNKMDVTEQLNATTFP----NVRIAARKQGSKDILRLRLIVDEIRHIFDE 361
 ++ L +YNK D + N P + ++A K+ LR++I + +F

40 Sbjct: 310 EVDQSQMLVVYNKAD---KPNLPPIPVHQNGIEMSAHKREDIQLRLQMIERTLVDLFTP 366

Query: 362 FSIRVHQNQAYKLYDLNKIALLDITYTFEEYE--NITGYISPKQKW 405
 + + ++ KL L + ++ ++E+ E + GY+ P W

40 Sbjct: 367 YVTELASDEGNKLAKLRRETIMTKWDEDECYQVKGYPVHPNHAW 412

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2615> which encodes the amino acid sequence <SEQ ID 2616>. Analysis of this protein sequence reveals the following:

Possible site: 48

>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----

45 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 50 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

55 >GP:BAB06081 GB:AP001515 unknown conserved protein [Bacillus halodurans]
 Identities = 185/403 (45%), Positives = 246/403 (60%), Gaps = 6/403 (1%)

Query: 13 ERVILLGVEL--QTTEHFDMSMTELANLAKTAGVKVMASFSQKRERYDSKTFISGKLDE 70
 ERV L+ +L T E F+ S+ EL L TA V+ +QKRE + T+IG GKLD

60 Sbjct: 10 ERVFLVACQLPNMTDEQFEASLEELALTLTAQGTVIDRLTQKREAIEPATYIGRGKLDE 69

Query: 71 IKAIVEADEIDAVIVNNRLTARQANLEAVLEVKVIDRMQLILDIFAMRARSHEGKLQVH 130
 + +E E D VIVN L+ Q NL L V+VIDR QLILDIFA RA+S EGKLQV

60 Sbjct: 70 LAIKMEEQEADLVIVNGELSGSQVRNLTNRLGVRVIDRTQLILDIFAGRAKSREGKLQVE 129

Query: 131 LAQLKYMLPRLVGQGIMLSRQAGGIGSRGPESQLELNRRSIRHQIADIERQLTQVEKNR 190

-950-

LAQL Y+LPR+VGQG LSR GGIG+RGPGE++LE +RR IR ++ADI++QL K+R
 Sbjct: 130 LAQLNYLLPRIVGQGQSLRGGIGTRGPGETKLETDRRHIRKRMADIDKQLKHTVKHR 189

Query: 191 QTIRDRRVGSDTFKIGLIGYTNAGKSTIMNLLTDDSHYEANELFATLDATTKQLYLENQF 250
 R RR + TF+I L+GYTNAGKST++N LT YE + LFATLD T+++ L +
 Sbjct: 190 DRYRARRERNQTFRIALVGYNAGKSTLLNRLTASDSYEEDLLFATLDEMTKMRPLPSGM 249

Query: 251 QATLTDTVGFIQDLPTLVAAPKSTLEESKYVDLLHVIDASDPNHSEKQKVVINLLKEL 310
 + L+DTVGF LPT LVAAF+STLEE K+ DLLLHV+D S + V LL +L
 Sbjct: 250 EVILSDTVGFINQLPTTLVAAPKSTLEEVKHADLLHVVDRSSEQLQAHMETVSELLHQL 309

Query: 311 DMLNIPRLAIYNKVDIAEQ--FTATAFPNIRISARSKSKILLRRLIIDQIRDQFVPFRI 368
 ++ L +YNK D I +SA ++ LR++I + D F P+
 Sbjct: 310 EVDQSQMLVVYNKADKPNLPPIFVHQNGIEMSAHKREDIQLRQMIERTLVLDLFTPYVT 369

Query: 369 KVHQDKAYKLYDLNRVALLDHYTFDQEIE--DISGYISPKQW 409
 ++ D+ KL L R ++ +D++ E + GY+ P W
 Sbjct: 370 ELASDEGNKLAKLRRETIMTEMKWDEDECYQVKGYVHPNHAW 412

20 An alignment of the GAS and GBS proteins is shown below.

Identities = 326/412 (79%), Positives = 375/412 (90%)

Query: 1 MIETKEEQERVILVGVELQDTENFEMSMEELASLAKTAGANVVNHYYQKRDKYDSKSF 60
 MIETK +QERVIL+GVELQ TE+F+MSM ELA+LAKTAG V+ + QKR++YDSK+FIG
 Sbjct: 5 MIETKRQQERVILVGVELQDTTEHFDMSMTLANLAKTAGVKVMASFSQKRERYDSKTF 64

Query: 61 SGKLEIKAIVEADEIDTVVNNRLTPRQNSNLEAELGVKVIDRMQLILDIFAMRARSHE 120
 SGKL+EIKAIVEADEID V+VNNRLT RQN+NLEA L VKVIDRMQLILDIFAMRARSHE
 Sbjct: 65 SGKLDEIKAIVEADEIDAVIVNNRLTARQANLEAVLEVKVIDRMQLILDIFAMRARSHE 124

Query: 121 GKLVHQAQLKYMFLPRLVGQGIIMLSRQAGGIGSRGPGEQLELNRRSIRHQISDIERQLK 180
 GKLVHQAQLKYMFLPRLVGQGIIMLSRQAGGIGSRGPGEQLELNRRSIRHQI+DIERQL
 Sbjct: 125 GKLVHQAQLKYMFLPRLVGQGIIMLSRQAGGIGSRGPGEQLELNRRSIRHQIADIERQLT 184

Query: 181 IVEKNRETVRERRVDSTTFKIGLIGYTNAGKSTIMNVLTDDKQYEANELFATLDATTKQI 240
 VEKNR+T+R+RRV S TFKIGLIGYTNAGKSTIMN+LTDD YEANELFATLDATTKQ+
 Sbjct: 185 QVEKNRQTIRDRRVGSDTFKIGLIGYTNAGKSTIMNLLTDDSHYEANELFATLDATTKQL 244

Query: 241 YLQNQFQVTLTDTVGFQDLPTLVAAPKSTLEESRHVDLLFHVIDASDPNHEEHEKVVM 300
 YL+NQFQ TLTDTVGFQDLPTLVAAPKSTLEES++VDLL HVIDASDPNH E EKVV+
 Sbjct: 245 YLENQFQATLTDTVGFIQDLPTLVAAPKSTLEESKYVDLLHVIDASDPNHSEKQKVV 304

Query: 301 EILKDLMDIPRLAIYNKMDVTEQLNATTFPNVRIAACKQGSKDLLRRLIVDEIRHIFD 360
 +LK+LDM++IPRLAIYNK+D+ EQ AT FPN+RI+A+ + SK LLRLI+D+IR F
 Sbjct: 305 NLLKELMDLNIPLAIYNKVDIAEQFTATAFPNIRISARSKSKILLRRLIIDQIRDQFV 364

Query: 361 EFSIRVHQNAQAYKLYDLNKLALDITYTFEEYENITGYISPKQKWKLEEFYD 412
 F I+VHQ++AYKLYDLN++ALLD YTF++E E+I+GYISPKQ+W+L++FY+
 Sbjct: 365 PFRIKVHQDKAYKLYDLNRVALLDHYTFDQEIEDISGYISPKQWRLDDFY 416

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 864

A DNA sequence (GBSx0916) was identified in *S.agalactiae* <SEQ ID 2617> which encodes the amino acid sequence <SEQ ID 2618>. Analysis of this protein sequence reveals the following:

Possible site: 46

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

60 bacterial cytoplasm --- Certainty=0.2044(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2619> which encodes the amino acid sequence <SEQ ID 2620>. Analysis of this protein sequence reveals the following:

```

5   Possible site: 40
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
10      bacterial cytoplasm --- Certainty=0.3436(Affirmative) < succ>
        bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

15      Identities = 124/209 (59%), Positives = 150/209 (71%)

    Query: 1  MIDYIDLALTYGGFTSLDKVYLEKKLDGLSKQQRLDFITPPPSVINAYFAEIIYQKQGPEA 60
              M +YIDLA TYGGFTSLD YL L L+ QQ+L FITPPPSVINAYFAEIIYQKQ P+A
    Sbjct: 5  MNNYIDLAKTYGGFTSLDTNYLNHLLASLTDQQLAFITPPPSVINAYFAEIIYQKQSPQA 64

20      Query: 61 ATDYFDFLSKALGLFPKHLSDDEEKPFIRLNLSGKSGFGAYLNDQEEASVFSEVKEVITP 120
                  ATDYF+LSKALGLF SF+EEKPF+RLNLSGK++GFAY NDQE A VFSE E P
    Sbjct: 65 ATDYFNLKALGLFTDQPSFEEKPFVRLNLSGKAYGFAYQNDQEQVALVFSEKAEPPKP 124

25      Query: 121 QLLLEIAQIFPQYKVYRDRSGIRMAKIDFDETESQNITPETSLLGNVLQLKKDIIKITSF 180
                  +L E+ QIFPQY VY D+ ++M F++ E ++ITP+ +LL + +L I + F
    Sbjct: 125 ELFFELTQIFPQYMYEDKGQLKMQAKQFEQGECEITPDDTLLSKIYRLANGITMLKGF 184

    Query: 181 NQEELLELVKTKSGKYYYSSQGRESVIYI 209
                  N EEL L +T SG+ YY RE +IYI
30      Sbjct: 185 NVEELWALSQTFSGQKYDFQAQREFMIYI 213

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 865

35 A DNA sequence (GBSx0917) was identified in *S.agalactiae* <SEQ ID 2621> which encodes the amino acid sequence <SEQ ID 2622>. Analysis of this protein sequence reveals the following:

```

    Possible site: 16
    >>> Seems to have no N-terminal signal sequence

40      ----- Final Results -----
        bacterial cytoplasm --- Certainty=0.1060(Affirmative) < succ>
        bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

45 A related GBS nucleic acid sequence <SEQ ID 9895> which encodes amino acid sequence <SEQ ID 9896> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

    >GP:CAB14316 GB:Z99116 similar to hypothetical proteins [Bacillus subtilis]
    Identities = 156/309 (50%), Positives = 210/309 (67%), Gaps = 5/309 (1%)

50      Query: 1  MEIQFLGTGAGQPAKARNVSSVLKLLDEINEVVMFDCGEGTQRILETTIKPRKVKKIF 60
                  ME+ FLGTGAG PAKARNV+S+ LKLL+E VW+FDCGE TQ QIL TTIKPRK++KIF
    Sbjct: 1  MELLFLGTGAGIPAKARNVTSVALKLEERRSVWLFDCEATQHQLHTTIKPRKIEKIF 60

55      Query: 61 ITHMHGDHVFGFLPGFLSSRAFAQNEEQTDLDIYGPVGIKSFVMTALRTSGSRLPYRIHFH 120
                  ITHMHGDHV+GLPG L SR+FQ E++ L +YGP GIK+F+ T+L + + L Y +

```

-952-

Sbjct: 61 ITHMHGDHVGLEPGLLSRSFQGGGEDE--LTVYGPKGKIAFIETSLAVTKTHLTYPLAIQ 118
 Query: 121 EFDESSLGKIMETDKFTVYAEKLDHTIFCMGYRVVQKDLEGLTDAEALKLAGVPFGPLFG 180
 E +E G + E D+F V A + H + GYRV +KD+ G+L A+ LK +P GP++
 5 Sbjct: 119 EIEE---GIVFEDDQFIVTAVSVIRGVGEAFGYRVQEKDVPGLSKADVLKEMNIPPGPVYQ 175
 Query: 181 KVKNGENVLTLEDGREIIAKDYISEPKKGKVTITILGDTTRKTDASIRLALGADVLVHESTYG 240
 K+K GE VTLEDGR I D++ PKKG+ + GDTR +D LA DVLVHE+T+
 10 Sbjct: 176 KIKKGETVTLEDGRIINGNDFLEPPKGRSVVFGSDTRVSDKLKELARDCDVLVHEATFA 235
 Query: 241 KGDERIAKSHGHSTNMQAADIAKQANAKRLLLNHVSARFMGRDCWQMEEDAKTIFSNTHL 300
 K D ++A + HST QAA AK+A AK+L+L H+SAR+ G +++++A +F N+
 Sbjct: 236 KEDRKLAYDYHSTTEQAAVTAKARAKQLILTHISARYQGDASLELQKEAVDVFPNSVA 295
 15 Query: 301 VRDLEEVGI 309
 D EV +
 Sbjct: 296 AYDFLEVNV 304

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2623> which encodes the amino acid
 20 sequence <SEQ ID 2624>. Analysis of this protein sequence reveals the following:

Possible site: 16
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 25 bacterial cytoplasm --- Certainty=0.2352(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

30 Identities = 253/307 (82%), Positives = 285/307 (92%)
 Query: 1 MEIQFLGTGAGQPAKARNVSSLVLKLLDEINEVWMFDCGEGTQRQILETTIKPRKVKKIF 60
 ME+QLGTGAGQPAK RNVSSL LKLLDEINEVWMFDCGEGTQRQILETTIKPRK++KIF
 Sbjct: 1 MELQFLGTGAGQPAQRNVSSLALKLLDEINEVWMFDCGEGTQRQILETTIKPRKIRKIF 60
 35 Query: 61 ITHMHGDHVFGLPGFLSSRAFAQNEEQTDLDIYGVPVGIKSFVMTALRTSGSRLPYRIHFH 120
 ITH+HGDH+FGFLPGFLSSR+FAQ+EEQTDLDIYGP+GIK++V+T+L+ SG+R+PY+IHFH
 Sbjct: 61 ITHLHGDHIFGLPGFLSSRSFAQSEEQTDLDIYGPIGIKTYVLTSLKVS GARVPYQIHFH 120
 40 Query: 121 EFDESSLGKIMETDKFTVYAEKLDHTIFCMGYRVVQKDLEGLTDAEALKLAGVPFGPLFG 180
 EFD+ SLGKIMETDKF VYAE+L HTIFCMGYRVVQKDLEGLTDAEALK AGVPFGPLFG
 Sbjct: 121 EFDDKSLGKIMETDKFEVYAEERLAHTIFCMGYRVVQKDLEGLTDAEALKAGVPFGPLFG 180
 45 Query: 181 KVKNGENVLTLEDGREIIAKDYISEPKKGKVTITILGDTTRKTDASIRLALGADVLVHESTYG 240
 K+KNG++V LEDGR I AKDYIS PKKGK+ITI+GDTRKT AS++LA ADVLVHESTYG
 Sbjct: 181 KIKNGQDVELEDGRLICAKDYISAPKKGKJITIIGDTRKTSASVKLAKDADVLVHESTYG 240
 Query: 241 KGDERIAKSHGHSTNMQAADIAKQANAKRLLLNHVSARFMGRDCWQMEEDAKTIFSNTHL 300
 KGDERIA++HGHSTNMQAA IA +A AKRLLLNHVSARF+GRDC QME+DA TIF N +
 50 Sbjct: 241 KGDERIARNHGHSTNMQAQIAHEAGAKRLLLNHVSARFLGRDCRQMEKDAATIFENVKM 300
 Query: 301 VRDLEEV 307
 V+DLEEV
 Sbjct: 301 VQDLEEV 307
 55

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 vaccines or diagnostics.

Example 866

A DNA sequence (GBSx0918) was identified in *S.agalactiae* <SEQ ID 2625> which encodes the amino acid sequence <SEQ ID 2626>. This protein is predicted to be similar to ketoacyl reductase. Analysis of this protein sequence reveals the following:

```

5   Possible site: 17
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
    bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
10   bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15   >GP:CAB14310 GB:Z99116 similar to ketoacyl reductase [Bacillus subtilis]
    Identities = 100/253 (39%), Positives = 152/253 (59%), Gaps = 2/253 (0%)

    Query: 3   RTILITGASGGLAQAIINQLPQDD-HLIVTGRSREKLEKLYGKRPNTLCLSLDITN-DNA 60
            + I ITGASGGL + I      + H++++ R ++L ++ K      +I D
    Sbjct: 7   KRIWITGASGGLGERIAYLCAAEGAHVLLSARREDRLIEIKRKITEWESGQCEIFPLDVG 66

20   Query: 61  VTNMIEKIYGERGQIDILINNAGFGSFKEFWDYSDDEEVKDMFAVNTFATMSIARQIGHKM 120
            I ++ + G ID+LINNAGFG F+ D + +++K MF VN F ++ + + +M
    Sbjct: 67  RLEDIARVRDQIGSIDVLINNAGFGIFETVLDSTLDDMKAMFDVNVFGLIACFKAVLPQM 126

25   Query: 121 SLVKSQGHIVNIASMAGLIATSKASVYGASKFAVVGFSNALRLELAENYVYVTSVNPQPIK 180
            K GHI+NIAS AG IAT K+S+Y A+K AV+G+SNALR+EL+ +YVT+VNPQPI+
    Sbjct: 127 LEQKKGHIINIASQAGKIATPKSSLYSATKHAVLGYSNALRMELSGTGIYVTTVNPQPIQ 186

30   Query: 181 TGFFAQADPSGDYLASIGRFALTPEKVSQKVVSVILGKNKRELNLPFILAFAHKYSLFPK 240
            T FF+ AD GDY ++GR+ L P+ V+ ++ + + KRE+NLP ++ K Y LFP
    Sbjct: 187 TDFFSIADKGGDYAKNVGRWMLDPPDVAQITAAIFTKKREINLPRLMNAGTKLYQLFPA 246

    Query: 241 TADYFARKVFENYK 253
            + A + K
35   Sbjct: 247 LVEKLAGRALMKK 259

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2627> which encodes the amino acid sequence <SEQ ID 2628>. Analysis of this protein sequence reveals the following:

```

40   Possible site: 18

    >>> Seems to have a cleavable N-term signal seq.

    ----- Final Results -----
    bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
45   bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

50   >GP:BA05225 GB:AP001512 oxidoreductase [Bacillus halodurans]
    Identities = 107/259 (41%), Positives = 156/259 (59%), Gaps = 5/259 (1%)

    Query: 1   MAQRIIVITGASGGLAQAIIVKQLPKEDSLI-LLGRNKERLEHCYQHI---DNKECLELD 55
            M ++ I ITGAS GL + +      E++++ L R++ERLE+ + +      +D
    Sbjct: 1   MRKKTIFITGASSGLGRQLAIDFSWEETVLCLFARSQERLENVQRIQVENGGEAHIYPVD 60

55   Query: 56  ITNPVAIEKMVAQIYQRYGRIDVLINNAGYGAFKGFEEFSAQEIADMVQVNTLASIHFAC 115
            + +P +I++ A+      G +DVLINNAGYG F+ F +      E MF+VN +
    Sbjct: 61  LADPQSIDRSFAEAIASVGVVDVLINNAGYGVFEFPFCDSDENERMFRVNVFGLMRATA 120

60   Query: 116 LIGQKMAEQGQHILINIVSMAGLIASAKSSIYSATKFALIGFSNALRLELADKGVYVTTV 175
            + M EQG GH+INI S AG IA+AKS+IYSATK A++GF+N+LR+EL G++V+ V

```

-954-

Sbjct: 121 AVLPTMREQSGHIINIASQAGKIATAKSAIYSATKHAVLGFTNSLRMELKGTGIHVS AV 180

Query: 176 NPGPIATKFFDQADPSGHYLESVGKFTLQPNQVAKRLVSIIGKNKRELNLFFSLAVTHQF 235
NPGPI T FFDQAD G Y V + L P V++++V + K KRELNLFP+ + +

5 Sbjct: 181 NPGPIQTFFDQADKEGAYTSKVQRIMLDPEDVSEKIVQLTKKPKRELNLFPWMMNIGATA 240

Query: 236 YTLFPKLS DY LARKVFNYK 254
Y + P+L + LA K F K

10 Sbjct: 241 YQVAPRLLELLAGKQFRQK 259

An alignment of the GAS and GBS proteins is shown below.

Identities = 155/251 (61%), Positives = 200/251 (78%)

15 Query: 3 RTILITGASGGLAQAIINQLPQDDHLIVTGRSREKLEKLYGKRPNTLCLSLDITNDNAV T 62
R I+ITGASGGLAQAI+ QLP++D LI+ GR++E+LE Y N CL LDITN A+

Sbjct: 4 RIIVITGASGGLAQAIIVQLPKEDSLILGRNKRERLEHCYQHIDNKECLELDITNPVAIE 63

Query: 63 NMIEKIYGBFGQIDILINNAGFGSFKEFWDSDEEVKDMFAVNTFATMSIARQIGHKMSL 122
M+ +IY +G+ID+LINNAG+G+FK F ++S +E+ DMF VNT A++ A IG KM+

20 Sbjct: 64 KMVAQIYQRYGRIDVLINNAGYGAFKGFEEFSAQEIADMPQVNTLASIHFACLIGQKMAE 123

Query: 123 VKSGHIVNIAS MAGLIATSKASVYGASKFAVVGFSNALRLELA EKNNVYVTSVNP GPIKTG 182
GH++NI SMAGLIA++K+S+Y A+KFA++GFSNALRLELA+K VYVT+VNPGPI T

25 Sbjct: 124 QGQGHILINIVSMAGLIASAKSSIYSATKFALIGFSNALRLELADKGVYVTVNPGPIATK 183

Query: 183 FFAQADPSGDY LASIGRFALTPEKVSKKVVSILGKNKRELNLFFILAF AHKYYS LF PKTA 242
FF QADPSG YL S+G+F L P +V+K++VSI+GKNKRELNLFF LA H++Y+LFPK +

Sbjct: 184 FFDQADPSGHYLESVGKFTLQPNQVAKRLVSIIGKNKRELNLFFSLAVTHQFYTLFPKLS 243

30 Query: 243 DYFARKVFNYK 253
DY ARKVFNYK

Sbjct: 244 DY LARKVFNYK 254

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 867

A DNA sequence (GBSx0919) was identified in *S. agalactiae* <SEQ ID 2629> which encodes the amino acid sequence <SEQ ID 2630>. This protein is predicted to be single-stranded-DNA-specific exonuclease (recJ). Analysis of this protein sequence reveals the following:

40 Possible site: 31
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -0.16 Transmembrane 197 - 213 (197 - 213)

45 ----- Final Results -----
bacterial membrane --- Certainty=0.1065(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

50 >GP:CAB14721 GB:Z99118 similar to single-strand DNA-specific
exonuclease [Bacillus subtilis]
Identities = 276/772 (35%), Positives = 447/772 (57%), Gaps = 45/772 (5%)

55 Query: 1 MISAKYSWVLNNQKPDAGFFFEASKKE-KISEAVASLIYSRGIKTS AELHHFLQTNLNLH 59
M+++K W + Q+PD ++ ++ I+ VASL+ RG T+ FL T + +

Sbjct: 1 MLASKMRWEI--QRFQDKVKSLTEQLHITPLVASLLVKRGFDTAESARLFLHTKDADPY 58

Query: 60 DPYLLNDMDKAVNRIRRAIENNETILVYGDYDADGMTSASIMKEALDMMGAEVQVYLPNR 119
DP+ + M +A +RI++AI E I++YGDYDADG+TS S+M L + A+V Y+P+R

60 Sbjct: 59 DPFEMKGMKEAADRIKQAISSQEKIMYGDYDADGVTSTSVMLHTLQKLSAQVDFYIPDR 118

-955-

5 Query: 120 FTDGYGPNQSVYKYFIEQQDVSIIITVDNGVAGHEAITYAQNQGVDDVVTDHHSMPADLP 179
F +GYGPN+ ++ I+++ SLIITVD G+A A+ G+DV++TDHH +LP
Sbjct: 119 FKEGYGPNQAFRS-IKERGFSLIITVDTGIAAVHEAKVAKELGLDVIITDHHEPGPELP 177

10 Query: 180 CAYAIHPEHPDANYPPFFYLAGCGVAFKVACALLETIPTMLDLVAIGTIADMVSLTDEN 239
AI+HP+ P YPF LAG GVAFK+A ALL +P E+LDL AIGTIAD+V L DEN
Sbjct: 178 DVRAIVHPKQPGCTYPFKELAGVGVAFLAHALLGELPDELDDLAAIGTIADLVPLHDEN 237

15 Query: 240 RIMVKAGLEVMDKSERIGLQELISLSNIDLKTNEETIGFKIAPQLNALGRLDDPNPAIE 299
R++ GLE ++ + R+GL+ELI LS D+ NEET+GF+++AP+LNA+GR++ +PA+
Sbjct: 238 RLIAITLGLERLRRTNRLGLKELIKLSSGGDIGEANEETVGFQIAPRLNAVGRIEQADPAVH 297

20 Query: 300 LLTGFDDEESQAIAQMIDQKNEERKEIVQTIIFDQAMQMLDQ---TKPVQVLAKENWHFPGV 356
LL D E++ +A IDQ N+ER+++V + D+A++M++Q + V+AK W+PGV
Sbjct: 298 LLMSSEDSFEAEELAAEIDQLNKRQKMVSKMTDEATEMVEQQGLDQTAIVVAKAGWNEPGV 357

25 Query: 357 LGIVAGRILERTGQPVIVLNI--EDGIAKGSARSVEALDIFQAFDQHRELFIAPFGHSGA 414
+GIVA ++++R +P IVL I E GIAKGSARS+ ++F++ + R++ FGGH A
Sbjct: 358 VGIVASKLVDRFYRPAIVLGLDEEKGIKGSARSIRGFNLFSLSSECRDILPHFGGHPMA 417

30 Query: 415 AGMTLEESKVGDLSQLCDYISKKQLDMSQKTLTIDSELRFDLSLDTVRDFEKLAPFG 474
AGMTL+ V DL L + + +D +++++++ + L+FFG
Sbjct: 418 AGMTLKAEDVPDLRSRLNEIADNTLTEDFIPQVEVDLVCGVEDITVESIAEMNMLSPFG 477

35 Query: 475 MDNKKPVFLLKDFKVSQARVMQNGAHLKLEQDQALDLVAFNMGSQLEFQQAQHL 534
M N KP L+++ + R +G N H+K+ + + LD V FN G + +
Sbjct: 478 MLNPKPHVLVENAVLEDVRKIGANKTHVKMTIRNESSQLDCVGFNKGELQEGIVPGSRIS 537

40 Query: 535 LAVTLSVNOWNGATTQLQMLEARVDGIQLFDIRSK-----ASSLPHG----- 577
+ +S+N+WN QLM++DA V QLFD+R K S+LP
Sbjct: 538 IVGEMSINENNRKKPQLMIKDAVSEWQLFDLRGKRTWEDTVSALPSAKRAIVSFKEDS 597

45 Query: 578 -----VPILSQEERQSK-----VILLTVPDHPQELKQMTQKQFDAIYFKN 618
V ++S ++Q+K ++LL P L ++ +GK + IYF
Sbjct: 598 TTLLQTEDLRREHVHVISSKQAKAFDLDGAYIVLLDPPPSLDMLARLLEKAPERYIFIF 657

50 Query: 619 EIPKNYFISGYGTRDQFASLYKTIYQFPEFDVRYKLKELSSYLHIPPILLIKMIQIFEE 678
+++F+S + RD F Y + + FDV+ EL+ + + M ++F +L
Sbjct: 658 LNHEDHFLSTFPARDHFQWYVAFLLKRGAFDVKKHGSELAKHKGWSVETINFMTKVFFDL 717

55 Query: 679 HFVTITEGIMTVNKEAEKRDISESQIYQELKETVKFQELMALGTPKEIYDFM 730
FV I G+++V A+KRD+++SQ YQ ++ ++ + + +E+ +++
Sbjct: 718 GFVKIENGVLSSVSGAKKRDLTDSQTYQAKQQLMELDQKLNYSAAEELKEWL 769

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2631> which encodes the amino acid sequence <SEQ ID 2632>. Analysis of this protein sequence reveals the following:

Possible site: 31
>>> Seems to have no N-terminal signal sequence

50 INTEGRAL Likelihood = -0.16 Transmembrane 220 - 236 (220 - 236)
INTEGRAL Likelihood = -0.11 Transmembrane 667 - 683 (667 - 683)

----- Final Results -----

55 bacterial membrane --- Certainty=0.1065(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 474/731 (64%), Positives = 594/731 (80%)

60 Query: 1 MISAKYSWVLNNQKPDAGFFFEASKKEKISEAVASLIYSGIKTSARELHHFLQTNLENLHD 60
MI +KYSW + ++KPD GFF+ +K + +++ A LIY RGI+T L FL +L LHD
Sbjct: 1 MIKSKYSWKIKDKKPDGFFFLAKTKGLTQTAAQLIYDRGIRTEALDEFLTADLSQLHD 60

65 Query: 61 PYLLNDMDKAVNRIRRAIENNETILVYGDYDADGMTSASIMKEALDMMGAEVQVYLPNRF 120

-956-

PYLL+DM KAV RIR+ATE E IL+YGDYDADGMTSASI+KE LDMMGAE VYLPNRF
 Sbjct: 61 PYLLHDMKAVPRIRQAIEBGERILIYGDYDADGMTSASIVKETLDMMGAEPLVYLPNRF 120
 Query: 121 TDGYGPNQSVYKYFIEQQDVSLIITVDNGVAGHEAITYAQNQGVDDVVTDHHSMPADLPC 180
 TDGYGPNQSVYKYFIEQ+ VSLIITVDNGVAGHEAI YAQ Q VDV+VTDHHS+P +LP
 Sbjct: 121 TDGYGPNQSVYKYFIEQEA VSLIITVDNGVAGHEAIRYAQEQEVDVIVTDHHSLEPEELPE 180
 Query: 181 AYAIHPEHPDANYPPPYLAGCGVAFKACALLETIPTMLDLVAIGTIADMVSLTDENR 240
 A+AIHPEHPDA+YPF +LAGCGVAFK+A ALLE++PT+ LDLVAIGTIADMVSLT ENR
 Sbjct: 181 AFAIHPDADYPFKHLAGCGVAFKLALLESLESLPTDCLDLVAIGTIADMVSLTGENR 240
 Query: 241 IMVKAGLEVMDKSERIGLQELISLSNIDLKITLNEETIGFKIAPQLNALGRLLDDPNPAIEL 300
 ++VK GL ++K +ER+GLQEL+SLS IDL+ NE+ IGF+IAPQLNALGRLLDDPNPAIEL
 Sbjct: 241 VLVKNGLAMLKHTERVGLQELMSLSPIDLEHFNEDAIGFQIAPQLNALGRLLDDPNPAIEL 300
 Query: 301 LTGFDDDEESQAIQMIQDKNEERKEIVQTFDQAMQMLDQTKPVQVLAKENWHHPGVLGIV 360
 LTGFDD+E+QAIA MI +KNEERK +VQ IFDQAM M+D KPVQVLA+ WHHPGVLGIV
 Sbjct: 301 LTGFDDQEAQAIQMIKKNEERKALVQDIFDQAMAMVDPQKPVQVLAQAGWHHPGVLGIV 360
 Query: 361 AGRILERTGQPVIVLNIEDGIAKGSARSVEALDIFQAFDQHRELFIAFGGHSGAAGMTLE 420
 AGRI+E QQ V+VL I++G AKGSARS+EA++IF+A + RELF AFGGH+GAAGMTL
 Sbjct: 361 AGRIMETIGQTVVLTIDNGFAKGSARSLEAINIFEALNGKRELFTAFGGHAGAAGMTLP 420
 Query: 421 ESKVGDLSQLCDYISKQDMSQKTLTIDSELRFDELSLDTVRDFEKLAPFGMDNKKP 480
 + LS LC ++ ++ LD + K TLTID L D+LSLD ++ +KLAP+GMD++KP
 Sbjct: 421 VDNLEALSDFLCQFVIERGLDQTAKNTLTIDERLSLDDLSDLILKSLDKLAPYGMHQK 480
 Query: 481 VFLLKDKFVSQARVMGQNGAHLKLEQDQALDLVAFNMGSQLEFQQAQHLAVTIS 540
 VF +KD +VSQAR +GQ+ +HLK K+ Q + D++AF GSQLQEF+QA LELAVTIS
 Sbjct: 481 VFYVDIRVSQARTIGQDQSHLKFVSQKASFDVLAFGQGSQLEFRQATGLELAVTIS 540
 Query: 541 VNQWNGATTQLMLEDARVDGIQLFDIRSKASSLPHGVPILSQEEQSKEVILLITVPDHPQ 600
 VN WNG T+LQ ML DARVDG+QL D+R+K + +P G+P + ++ ++ +++ +P+ +
 Sbjct: 541 VNHWNNTSLQFMLVDARVDGVLQLDLRTKTAKVPEGIPTIEEDPNARVILINDIPEDFK 600
 Query: 601 ELKQMTQGGQFDIYFKNEIPKNYFISGYGTRDQFASLYKTIYQFPEFDVRYKLKELSSY 660
 + K FDIYFKN++ Y+++G+G+R+QFA LYKTIYQFPEFD+R+KL ELS Y
 Sbjct: 601 TWRNQFVHKDFDIYFKNQMKHPYLTGFGSREQFAKLYKTIYQFPEFDLRHKLTELSHY 660
 Query: 661 LHIPDILLIKMIQIFEEELHFVTITEGIMTVNKEAKRDISESQIYQELKETVVKFQELMAL 720
 L+I +LLIK+IQIFEEEL FVTI +G+MTVN +A+KR+ISES IYQ+LKE VKFQE+MAL
 Sbjct: 661 LNIKLLLIKLIQIFEEELS FVTIDDLGMLTVNPQAQKREISESHIYQDLKELVKFQEIMAL 720
 Query: 721 GTPKEIYDFMM 731
 +PKE+YD+++
 Sbjct: 721 ASPKEMYDYLV 731

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

50 Example 868

A DNA sequence (GBSx0920) was identified in *Sagalactiae* <SEQ ID 2633> which encodes the amino acid sequence <SEQ ID 2634>. Analysis of this protein sequence reveals the following:

Possible site: 13
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4114(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

-957-

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 869

- 5 A DNA sequence (GBSx0921) was identified in *S.agalactiae* <SEQ ID 2635> which encodes the amino acid sequence <SEQ ID 2636>. Analysis of this protein sequence reveals the following:

```

Possible site: 42
>>> Seems to have an uncleavable N-term signal seq
    INTEGRAL    Likelihood = -5.10    Transmembrane    15 - 31 ( 14 - 33)
10
----- Final Results -----
        bacterial membrane --- Certainty=0.3039(Affirmative) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>
15        bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAA88584 GB:M18954 fructosyltransferase [Streptococcus mutans]
Identities = 67/219 (30%), Positives = 106/219 (47%), Gaps = 31/219 (14%)
20
Query: 1  MRPIVRKKMYKKKGKFWVAGIVT-ILGGSAILGQDVKAQEAEAVTSTISEKTDSSQTISD 59
      M  VRKKMYKKKGKFWVVA I T +L G +   V+A++A + T  SE + SQ +
Sbjct: 1  METKVRKKMYKKKGKFWVATITTTAMLTGIGL--SSVQADEANS-TQVSSELAERSQVQEN 57

Query: 60  TSKLTLPVNSSEAMKNSAEPLIKTAFATSVSSNPRIEAAATPVKTFDASSKVVKASTAEH 119
      T+      SS A +N A  KT  + S+NP  AA V+  D ++KV+  + E
25 Sbjct: 58  TTA-----SSSAAENQA----KTEVQETPSTNP---AAATVENTDQTTKVITDNAAVES 104

Query: 120  SANQTN---SNVNQVANDSEVITQQN-----STKQLPTVTYSAHVQDIGW----QKSVD 166
      A++T   + V + A + + Q N   +TK+   T   + + G   +K
30 Sbjct: 105 KASKTKDQAATVTKTAASTPEVGQTNEKDKAKATKEADITTPKNTIDEYGLTEQARKIAT 164

Query: 167  NATVSGTVGQEKQVEAIKLSIKAPEGITG-KLSYKTYVK 204
      A ++ +   +KQVEA+   +   TG +++Y+ + K
35 Sbjct: 165 EAGINLSSLTQKQVEALNKVKLTSDAQTGHQMTYQEFDK 203

```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

- 40 A related GBS gene <SEQ ID 8677> and protein <SEQ ID 8678> were also identified. Analysis of this protein sequence reveals the following:

```

Lipop Possible site: -1    Crend: 5
McG: Discrim Score:      9.08
GvH: Signal Score (-7.5): -3.94
Possible site: 34
45 >>> Seems to have an uncleavable N-term signal seq
    ALOM program count: 1 value: -5.10 threshold: 0.0
    INTEGRAL    Likelihood = -5.10    Transmembrane    7 - 23 ( 6 - 25)
    PERIPHERAL  Likelihood = 4.03     694
50 modified ALOM score: 1.52

*** Reasoning Step: 3

----- Final Results -----
        bacterial membrane --- Certainty=0.3039(Affirmative) < succ>
55        bacterial outside --- Certainty=0.0000(Not Clear) < succ>
        bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

Streptococcus mutans

SEQ ID 8678 (GBS243) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 57 (lane 7; MW 94kDa).

GBS243-His was purified as shown in Figure 208, lane 10.

Example 870

A DNA sequence (GBSx0922) was identified in *S.agalactiae* <SEQ ID 2637> which encodes the amino acid sequence <SEQ ID 2638>. This protein is predicted to be adenine phosphoribosyltransferase (apt).

5 Analysis of this protein sequence reveals the following:

Possible site: 59

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -1.86 Transmembrane 61 - 77 (59 - 77)

INTEGRAL Likelihood = -0.64 Transmembrane 137 - 153 (137 - 153)

10

----- Final Results -----

bacterial membrane --- Certainty=0.1744(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

15

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC46040 GB:U86377 adenine phosphoribosyltransferase; Apt

[Bacillus subtilis]

Identities = 110/170 (64%), Positives = 135/170 (78%)

20

Query: 1 MDLNNYIASIENYPQEGITFRDISPLMADGKAYSAYVREIVQYAADKDIDMIVGPEARGF 60

MDL Y+ + +YP+EG+ F+DI+ LM G Y YA +IV+YA +K ID++VGPEARGF

Sbjct: 1 MDLKQYVTVIPDYPKEGVQFKDITTLMDKGDVYRYATDQIVEYAKEQIDLVVGPEARGF 60

25

Query: 61 IVGCPVAYALGIGFAPVRKPKGLPREVISADYEKEYGLDITLTHADAIPGQQRVLIVDDL 120

I+GCPVAYALG+GFAPVRK GKLPREVI DY EYG D LT+H DAIKPGQQRVLI DDL

Sbjct: 61 IIGCPVAYALGVGFAPVRKEGKLPREVIKVDYGLEYGKDVLTTHKDAIKPGQQRVLITDDL 120

30

Query: 121 LATGGTVKATIEKLGGLVAGCAFLVELDGLNGRKAIEGYDTKVLNMF 170

LATGGT++ATI+++E+LGGVVAG AFL+EL L+GR +E YD LM +

Sbjct: 121 LATGGTIEATIKLVEELGGVVAGIAFLIELSYLDGRNKLEDYDILTLMKY 170

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2639> which encodes the amino acid sequence <SEQ ID 2640>. Analysis of this protein sequence reveals the following:

35

Possible site: 40

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

bacterial outside --- Certainty= 0.300(Affirmative) < succ>

40

bacterial membrane --- Certainty= 0.000(Not Clear) < succ>

bacterial cytoplasm --- Certainty= 0.000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

!GB:Z99120 similar to opine catabolism [Bacillus sub... 231 1e-59

45

>GP:CAB15253 GB:Z99120 similar to opine catabolism [Bacillus subtilis]

Score = 231 bits (583), Expect = 1e-59

Identities = 138/363 (38%), Positives = 212/363 (58%), Gaps = 11/363 (3%)

50

Query: 5 IIGAGIVGSTAAYYLQSQGQKEVTIFDHGQ-GQATKAAAGIISPWFSKRRNKVWYRMARL 63

I+GAGI+G++ AY+L ++G + VT+ D + GQAT AAAGI+ PW S+RRN+ WY++A+

Sbjct: 6 IVGAGILGASTAYHLAKTGAR-VTVDRKEPGQATDAAAGIVCPWLSQRRNQDWYQLAKG 64

55

Query: 64 GADFYQQLINDLKEDGFATDFYQQNGIYVLKKQEELRDLVELALARKVESPIIGELAIAK 123

GA +Y+ LI+ L++DG + Y++ G + KL + E A R+ ++P IG++

Sbjct: 65 GARYYKDLIHQLEKDGESDTGYKRVGAISHTDASKLDKMEERAYKRREDAPETIGDITRL 124

Query: 124 NRKELGNDFKGLIGFDNCLYASGAARVEGAALCETLLKAS---GYPVIRQKVTLKQQG-- 178

-960-

+ E F L ++ SGAARV G ALC +LL A+ G VI+ +L +
 Sbjct: 125 SASETKKLFFILADGYESVHISGAARVNGRALCRSLLSAAEKRGATVIKGNASLLFENG 184
 Query: 179 -SGYEIAGHYF--DQVILAAGAWLPDLLRPLGYQVDVRPQKGQLLDYDVHHIISDTYPVV 235
 +G + F D VI+ AGAW ++L+PLG V QK Q++ +++ + ++PVV
 Sbjct: 185 VTGVQTDTKQFAADAVITVAGAWANEILKPLGIHFQVSFQKAQIMHFEMTDADTGSWFPV 244
 Query: 236 MPEGEIDLIPFNQGGKISVGTSHENDKGY-DLEPDWQVLKKLEMQALTYLPLLKEATQKTC 294
 MP + ++ F+ G+I G +HEND G DL ++ +AL P L +A
 Sbjct: 245 MPPSDQYILSFDNGRIVAGATHENDAGLDDLRTAGGQHEVLKALAVAPGLADAAVET 304
 Query: 295 RVGIRAYTSDYSPFYGVQVSGLNLYTASGLGSSGLTVGPLIGYELAQLLLGHEGLTTPSD 354
 RVG R +T + P G V ++ LY A+GLG+SGLT+GP +G ELA+L+LG + L S
 Sbjct: 305 RVGFRPFTPGFLPVVGAVPNVQGLYAANGLGASGLTMGPFLGAELAKLVLGKQTELDLSP 364
 Query: 355 YSP 357
 Y P
 Sbjct: 365 YDP 367

20 An alignment of the GAS and GBS proteins is shown below.

Identities = 150/172 (87%), Positives = 161/172 (93%)

Query: 1 MDLNNYIASIENYPQEGITFRDISPLMADGKAYSAYAVREIVQYAADKDIDMIVGPEARGF 60
 MDL NYIASI++YP+ GITFRDISPLMADGKAYSAY+REI QYA DKDIDM+VGPEARGF
 Sbjct: 1 MDLTNYIASIKDYPKAGITFRDISPLMADGKAYSAYAREIAQYACDKDIDMVVGPEARGF 60
 Query: 61 IVGCPVAYALGIGFAPVRKPGKLPREVISADYEKEYGLDTLTMHADAIKPGQRVLIIVDDL 120
 I+GCPVA LGIGFAPVRKPGKLP+V+SADYEKEYGLDTLTMHADAIKPGQRVLIIVDDL
 Sbjct: 61 IIGCPVAVELGIGFAPVRKPGKLP+V+SADYEKEYGLDTLTMHADAIKPGQRVLIIVDDL 120
 Query: 121 LATGGTVKATIEKIEKLGIVAGCAFLVELDGLNGRKAIEGYDTKVLMMNFPF 172
 LATGGTVKATIEKIEKLG+VAGCAFL+EL+GINR AI YD KVLMM FPG
 Sbjct: 121 LATGGTVKATIEKIEKLGIVAGCAFLIELEGLNGRHAIRNYDYKVLMMQFPF 172

35 SEQ ID 2638 (GBS419) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 79 (lane 6; MW 22.6kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 172 (lane 4; MW 47.5kDa).

GBS419-GST was purified as shown in Figure 219, lane 6-8.

40 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 871

A DNA sequence (GBSx0923) was identified in *S.galactiae* <SEQ ID 2641> which encodes the amino acid sequence <SEQ ID 2642>. Analysis of this protein sequence reveals the following:

45 Possible site: 29
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0847(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 50 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

55 >GP:BAA11244 GB:D78182 ORF2 [Streptococcus mutans]
 Identities = 140/225 (62%), Positives = 178/225 (78%)
 Query: 1 MTYLEQYQSQQLTLPALFFHFKSIFKTADDFLVWQFFYLQNTITNLSDLTPSRIATSLDK 60
 M++L+ Y+SG L LPSAL FH+K IF ADDFLVWQFFY QNTIT + D+ S+IAT++ K

-961-

Sbjct: 1 MSFLQHYKSGNLVLP SALLFHYKDIFSNADDFLVWQFFYFQNTTKMEDIATSQIATAIGK 60

Query: 61 TVADINRSISNLTSQGLLDVKTIELNHEIEIIFDTS PVFAKLDKLFEBDNQVIIDNKTS D 120
TV ++NRS+SNL SQ LLD+KTIEL+ E E++FD + KLD L ++ + +

5 Sbjct: 61 TVPEVNRSVSNLISQELLDKMTIELDGESEVLF DATLALKKLDLLTADETTVSSSKGT 120

Query: 121 SNRLKDLVGDFERELGRLLSPFELEDLQKTLQEDQTD PDIVRAALREAVFNGKTSWNYIN 180
SN LKDLV DFERELGR+LSPFELEDLQKT+ +D+TDPD+VR+ALREAVFNGKT+WN YI

10 Sbjct: 121 SNALKDLVEDFERELGRMLSPFELEDLQKTVSDDKTD PDLVRSALREAVFNGKTNWNYIQ 180

Query: 181 AILRNWRREGLTTLRQIEERKQAREDNQMKDLAISDDFKNAMNLW 225
AILRNWRREG++TLRQ+EER++ RE ++ +SDDF +AMNLW

Sbjct: 181 AILRNWRREGISTLRQVEERRKEREQANPANVTVSDDFLSAMNLW 225

- 15 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2643> which encodes the amino acid sequence <SEQ ID 2644>. Analysis of this protein sequence reveals the following:

Possible site: 57

20 >>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
25 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:BAAl1244 GB:D78182 ORF2 [Streptococcus mutans]
Identities = 154/228 (67%), Positives = 188/228 (81%), Gaps = 1/228 (0%)

30 Query: 1 MSFLEHYKSGNLVIPSALLFHYKDLFKSSDDFLVWQFFYLQNTTKRDDLAPSQIAHALGK 60
MSFL+HYKSGNLV+PSALLFHYKD+F ++DDFLVWQFFY QNTTK +D+A SQIA A+GK

Sbjct: 1 MSFLQHYKSGNLVLP SALLFHYKDIFSNADDFLVWQFFYFQNTTKMEDIATSQIATAIGK 60

35 Query: 61 SVADINKIISLTLNQGLDMRTIELTGEIEIIFDASPVIAKLDQLFVSQTATEIDKQE-T 119
+V ++N+ +S+L +Q LLD+TIEL GE E++FDA+ L KLD L + T + + T

Sbjct: 61 TVPEVNRSVSNLISQELLDKMTIELDGESEVLF DATLALKKLDLLTADETTVSSSKGT 120

Query: 120 PNHFKRLVDEFERELGRFLSPFELEDLEKTLRDDKTD PDLIREALKEAVFNGKTNWKYIQ 179
N K LV++FERELGR LSPFELEDL+KT+ DDKTDPD+R AL+EAVFNGKTNW YIQ

40 Sbjct: 121 SNALKDLVEDFERELGRMLSPFELEDLQKTVSDDKTD PDLVRSALREAVFNGKTNWNYIQ 180

Query: 180 AILRNWRKEGIVNLRQVEERRRVREGEDLSQVTISEDFLSAMNLWSDS 227
AILRNWR+EGI LRQVEERR+ RE + + VT+S+DFLSAMNLWSDS

45 Sbjct: 181 AILRNWRREGISTLRQVEERRKEREQANPANVTVSDDFLSAMNLWSDS 228

An alignment of the GAS and GBS proteins is shown below.

Identities = 144/225 (64%), Positives = 179/225 (79%), Gaps = 1/225 (0%)

50 Query: 1 MTYLEQYQSGQLTLPSALFFHFKSIFKTADDFLVWQFFYLQNTTNLSDLTPSRIATSLDK 60
M++LE Y+SG L +PSAL FH+K +FK++DDFLVWQFFYLQNTT DL PS+IA +L K

Sbjct: 1 MSFLEHYKSGNLVIPSALLFHYKDLFKSSDDFLVWQFFYLQNTTKRDDLAPSQIAHALGK 60

Query: 61 TVADINRSISNLTSQGLLDVKTIELNHEIEIIFDTS PVFAKLDKLFEBDNQVIIDNKTS D 120
+VADIN+ IS+LT+QGLLD++TIEL EIEIIFD SPV AKLD+LF ID K

55 Sbjct: 61 SVADINKIISLTLNQGLDMRTIELTGEIEIIFDASPVIAKLDQLFVSQTATEID-KQET 119

Query: 121 SNRLKDLVGDFERELGRLLSPFELEDLQKTLQEDQTD PDIVRAALREAVFNGKTSWNYIN 180
N K LV +FERELGR LSPFELEDL+KTL++D+TDPD++R AL+EAVFNGKT+W YI

60 Sbjct: 120 PNHFKRLVDEFERELGRFLSPFELEDLEKTLRDDKTD PDLIREALKEAVFNGKTNWKYIQ 179

Query: 181 AILRNWRREGLTTLRQIEERKQAREDNQMKDLAISDDFKNAMNLW 225
AILRNWR+EG+ LRQ+EER++ RE + + IS+DF +AMNLW

Sbjct: 180 AILRNWRKEGIVNLRQVEERRRVREGEDLSQVTISEDFLSAMNLW 224

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 872

A DNA sequence (GBSx0924) was identified in *S.agalactiae* <SEQ ID 2645> which encodes the amino acid sequence <SEQ ID 2646>. Analysis of this protein sequence reveals the following:

Possible site: 47
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

10 bacterial cytoplasm --- Certainty=0.1617(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

15 >GP:BAA11245 GB:D78182 ORF3 [Streptococcus mutans]
 Identities = 134/226 (59%), Positives = 170/226 (74%)

20 Query: 2 DLQLSKRLQKVANYVPKGARLLDVGSDHAYLPIFLLQMGYCDFAIAGEVVNGPYQSALKN 61
 ++ LS RLQ+VA++VPKGARLLDVGSDHAYLPI+LL+ G DFA+AGE++ GPY+SA+ N
 Sbjct: 7 EVSLSHRLQEVASFVPKGARLLDVGSDHAYLPIYLLQGLIDFAVAGEIIGPYESAVAN 66

25 Query: 62 VSEHGLTSKIDVRLANGLSAFEEADNIDTITICMGGRLIADILNNDIDKLQHVKTLLVLQ 121
 V+E GL+ +I VRLA+GL+A + D+ID ITICMGGRLIADIL DKL VK L+LQ
 Sbjct: 67 VNESGLSGQIAVRLADGLAALNDNDIDITICMGGRLIADILAAGSDKLN SVKQLILQ 126

30 Query: 122 PNNREDDLRLKWLAAANDFEIVAEDILTENDKRYEILVVKHGHMNLTAKEIRFGPFLLSNNT 181
 PNN EDDLRL WL ANDF I AE ++ + K YEILVV+ G + L+ K+LRFGPFL +
 Sbjct: 127 PNNCEDDLRLSWLVANDEMIKAEKVMKDRHKYIEILVVEKGKITLSKDLRFGPFLRQERS 186

35 Query: 182 TVFKEKWQNELNKLTFALNSIPNSKMEERAILEDKIQDIKEVLDES 227
 ++FKE+W+ EL KL AL +P K + L' KI+ I+EVL ES
 Sbjct: 187 SIFKERWRKELAKLELALTRVPAKKKADNMFLSTKIEQIREVLYES 232

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2647> which encodes the amino acid sequence <SEQ ID 2648>. Analysis of this protein sequence reveals the following:

Possible site: 47
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

40 bacterial cytoplasm --- Certainty=0.0803(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

45 Identities = 145/224 (64%), Positives = 173/224 (76%)

50 Query: 1 MDLQLSKRLQKVANYVPKGARLLDVGSDHAYLPIFLLQMGYCDFAIAGEVVNGPYQSALK 60
 MD QLS RL +VA YVPGK +LLDVGSDHAYLPIFL++ AIAGEVV GPY+SALK
 Sbjct: 1 MDSQLSNRLAQVAAYVPKGVKLLDVGSDHAYLPIFLVETNQISAAIAGEVVVRGPYESALK 60

55 Query: 61 NVSEHGLTSKIDVRLANGLSAFEEADNIDTITICMGGRLIADILNNDIDKLQHVKTLLVL 120
 NV++ GL I VRLANGL+AFEEAD++ ITICMGGRLIADIL +KLQ ++ LVL
 Sbjct: 61 NVTQSGLAHEHIQVRLANGLAAFEEDVTAITICMGGRLIADILEAGKEKLGIERLVL 120

 Query: 121 QPNNREDDLRLKWLAAANDFEIVAEDILTENDKRYEILVVKHGHMNLTAKEIRFGPFLLSNN 180
 QPNNREDDLRL WL+ N F+IVAE I+ ENDK YEI+V +HG L+A ELRFGP+L
 Sbjct: 121 QPNNREDDLRAWLSVNAFKIVAETIMAENDKYYEIIVAEHGEKALSATELRFGPYLSQEK 180

 Query: 181 TTVFKEKWQNELNKLTFALNSIPNSKMEERAILEDKIQDIKEVL 224

-963-

+ VFKEKWQ E++KL +AL+ IP K +ER +L KIQ IKEV+
 Sbjct: 181 SVVFKEKWQREMDKLAYALSCIPPEKTQERQLLLTKIQIKEVI 224

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 5 vaccines or diagnostics.

Example 873

A DNA sequence (GBSx0925) was identified in *S.agalactiae* <SEQ ID 2649> which encodes the amino acid sequence <SEQ ID 2650>. Analysis of this protein sequence reveals the following:

Possible site: 54
 10 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3245(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 15 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9893> which encodes amino acid sequence <SEQ ID 9894> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

20 >GP:BAA11246 GB:D78182 ORF4 [Streptococcus mutans]
 Identities = 187/262 (71%), Positives = 224/262 (85%)

Query: 2 MKARELIDVYETPCQELSMEGDISGLQIGSLDKEIKTVMVALDVRETTVAEAIERQVDL 61
 MKA ++I YE YCPQ+LS+EGDISGLQIG+LDKEIK +M+ALDVRETTVAEAI++VDL
 25 Sbjct: 1 MKASQIIKRYEAYCPQDLSLEGDISGLQIGTLDKEIKRLMIALDVRETTVAEAIKKVDL 60

Query: 62 LIVKHAPIFRPLKDLVATPQNKIYIDLLKSDIAVYVSHTNIDIVPGLNDWFCCELLDIQY 121
 LIVKHAPIFRPLK+LV T QN IY +L+K DIAVYVSHTNIDIVP+GLNDWFC+LLDI+
 Sbjct: 61 LIVKHAPIFRPLKNLVETAQNHIYFNLIKHDIAVYVSHTNIDIVPDGLNDWFCDLLDIKN 120

30 Query: 122 PDILSETSNYGIGRIGDIRPQSFEFFAWKIKDVFGLDVRLVSYDKSNPEIQRVAICGG 181
 ILS + + YGIGR+GDI P SFE A K+K +F LDSVRLVSY ++NP I R+AICGG
 Sbjct: 121 RRILSPSKDDYGIGRVGDISPLSFEDLAKVKKIFNLDVRLVSYGENNPLISRIAICGG 180

35 Query: 182 SGQSFYKEAIAKGADVFTVTDIYYHTAQEMITNGLLAIDPGHHIEVLVFSKIATMIEQWK 241
 SGQSFY+EA+ KGA V++TGDIYYHTAQEM+TNGLLA+DPGHHIEVLV K+A + W
 Sbjct: 181 SGQSFYQEALTGAQVYITGDIYYHTAQEMLTNGLLLALDPGHHIEVLVFRKLAEKFQIWS 240

40 Query: 242 LEKGWDISVLESKAPTNPFFYHM 263
 ++ WDI++LES+ TNFFYH+
 Sbjct: 241 CQENWDITILESQVNTNPFFYHL 262

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2651> which encodes the amino acid sequence <SEQ ID 2652>. Analysis of this protein sequence reveals the following:

45 Possible site: 53
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1804(Affirmative) < succ>
 50 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 169/262 (64%), Positives = 214/262 (81%)

55 Query: 2 MKARELIDVYETPCQELSMEGDISGLQIGSLDKEIKTVMVALDVRETTVAEAIERQVDL 61
 MKA+ LID YE +CP +LSMEGD+ GLQ+GSLDK+I+ VM+ LD+RE+TVAEAI+ +VDL

-964-

Sbjct: 3 MKAKTLIDAYEAFCLDLSMEGDVKGLQMSLDKDIRKVMITLDIRESTVAEAIKNEVDL 62

Query: 62 LIVKHAPIFRPLKDLVATPQNKIYIDLLKSDIAVYVSHTNIDIVPGLNDWFCCELLDIQY 121
+I KHAPIF+PLKDLV++PQ I +DL+K DI+VYVSHTNIDIVP GLNDWFC+LL+I+

5 Sbjct: 63 IITKHAPIFKPLKDLVSSPQRDILLDLVKHDISVYVSHTNIDIVPGLNDWFCDLLEIKE 122

Query: 122 PDILSETSNYGGIGRIGDIRPQSFEFFANKIKDVFGLDVRLVSYDKSNPEIQRVAICGG 181
LSET G+GIGRIG ++ Q+ E A K+K VF LD+VRL+ YDK NP I ++AICGG

10 Sbjct: 123 ATYLSETKEGFGIGRIGTVKEQALEELASKVKRVFDLDTVRLIRYDKENPLISKIAICGG 182

Query: 182 SGQSFYKEAIAKGADVFTGDIYYHTAQEMITNGLLAIDPGHHIEVLVFSKIATMIEQWK 241
SG FY++A+ KGADV++TGDIYYHTAQEM+T GL A+DPGHHIEVLV K+ ++WK

Sbjct: 183 SGGEFYQDAVQKGADVITGDIYYHTAQEMLTGELFAVDPGHHIEVLVFTKLEKLEKLGWK 242

15 Query: 242 LEKGWDISVLESKAPTNPFFYHM 263
E GWD+S++ SKA TNPF H+

Sbjct: 243 EENGWDVSISSKASTNPFSHL 264

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 874

A DNA sequence (GBSx0926) was identified in *S.galactiae* <SEQ ID 2653> which encodes the amino acid sequence <SEQ ID 2654>. This protein is predicted to be (). Analysis of this protein sequence reveals the following:

25 Possible site: 41
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
bacterial outside --- Certainty=0.3000(Affirmative) < succ>
30 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

35 >GP:CAB15253 GB:Z99120 similar to opine catabolism [Bacillus subtilis]
Identities = 148/368 (40%), Positives = 211/368 (57%), Gaps = 13/368 (3%)

Query: 1 MKKIAIIGAGAVGATLAYYSKEKDQVTVFDYGV-GQATKAAAGIISPWFSKRRNKAWY 59
MK I+GAG +GA+ AY+L+K +VTV D GQAT AAAGI+ PW S+RRN+ WY

40 Sbjct: 1 MKSYIIVGAGILGASTAYHLAKT-GARVTVDRKEPGQATDAAAGIVCPWLSQRRNQDWY 59

Query: 60 RMARLGADFYSKLVTDLQKDGFTKPYQQTGVFLLKKDESQLESFLALADKRRLESPLIG 119
++A+ GA +Y L+ L+KDG Y++ G + D S+L+ + A KRR ++P IG

Sbjct: 60 QLAKGARYYKDLIHQLEKDGESDTGYKRVGAISIHTDASKLQDKMERAYKRREDAPRIG 119

45 Query: 120 DLQILNKSEANTHFPPEL-DGYEQLLYASGGARVEGADLTRILLEAS---GVNVIKDEVHF 175
D+ L+ SE FP L DGYE ++ SG ARV G L R LL A+ G VIK

Sbjct: 120 DITRLSASETKKLPILADGYES-VHISGAARVNGRALCRSLLSAAEKGATVIKGNASL 178

50 Query: 176 -----TITDNGFRVQGIIDFKLVLASGAWLAKILDEHNYQVDVRPQKQLRDYYFSNINT 230
T+T + D +++ +GAW +IL V QK Q+ + ++ +T

Sbjct: 179 LFENGTVTVGVQTDTKQFAADAVIVTAGAWANEILKPLGIHFQVSFQKAQIMHFEMTDADT 238

Query: 231 GKYPVVMPEGELDIIPFDNGKVSVGASHENDMAF-DLNIDFKVLDFEEOAIGYFPQLKK 289
G +PVVMP + I+ FDNG++ GA+HEND DL + + +A+ P L

55 Sbjct: 239 GSWPVVMPPSDQYILSFDNGRIVAGATHENDAGLDDLRTVAGGQHEVLKALAVAPGLAD 298

Query: 290 ADTTSERVIGIRAYTSDFSPFFGPVPCMEGAYAASGLGSTGLTVGPLIGYELCQLILNKEN 349
A RVG R +T F P G VP ++G YAA+GLG++GLT+GP +G EL +L+L K+

60 Sbjct: 299 AAAVETRVGFRPFTFGFLPVVGAVPNVQGLYAANGLGASGLTMGPFLGAEALAKNLVIGKQT 358

Query: 350 QLNLEDYD 357

-965-

+L+L YD
Sbjct: 359 ELDLSPYD 366

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2655> which encodes the amino acid
5 sequence <SEQ ID 2656>. Analysis of this protein sequence reveals the following:

Possible site: 40
>>> Seems to have a cleavable N-term signal seq.

10 ----- Final Results -----
bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

15 Identities = 211/360 (58%), Positives = 262/360 (72%)
Query: 3 KIAIIGAGAVGATLAYYLSKEKDIQVTVFYGVGQATKAAAGIISPWFSKRRNKAWYRMA 62
KIAIIGAG VG+T AYYL + +VT+FD+G GQATKAAAGIISPWFSKRRNK WYRMA
20 Sbjct: 2 KIAIIGAGIVGSTAAYYLQSQGQKEVTIFDHGQGQATKAAAGIISPWFSKRRNKVWYRMA 61
Query: 63 RLGADFYSKLVTDLQKDGFEFKFYQQTGVFLLKKDESQLESFALADKRRLESPLIGDLQ 122
RLGADFY +L+ DL++DGF T FYQQ G+++LKK E +L L+ LA R++ESP+IG+L
Sbjct: 62 RLGADFYQQLINDLKEDGFATDFYQQNGIYVLKKQEEKLRDLYELALARKVESPIIGELA 121
25 Query: 123 IILNKSEANTHFPPELDGYEQLLYASGGARVEGADLTRILLEASGVNVIKDEVHFTITDNGF 182
I N+ E F L G++ LYASG ARVEGA L LL+ASG VI+ +V +G+
Sbjct: 122 IKNRKELGNDFKGLIGFDNCLYASGAARVEGAALCETLLKASGYPVIRQKVTLLKQQSGY 181
30 Query: 183 RVQGIDFDKLVLASGAWLAKILDEHNYQVDVRPQKGQLRDYFYSNINTGKYPVVMPEGEL 242
+ G FD+++LA+GAWL +L YQVDVRPQKGQL DY +I + YPVVMPEGE+
Sbjct: 182 EIAGHYFDQVILAAGAWLPDLRLPLGYQVDVRPQKGQLLDYDVHHIISD'TYPVVMPEGEI 241
Query: 243 DIIPFDNGKVSVGASHENDMAFDLNDKFKVLDKFEEQAIGYFFQLKKADTTSERVGIRAY 302
D+IPF+ GK+SVG SHEND +DL D++VL K E QA+ Y P LK+A + RVGIRAY
35 Sbjct: 242 DLIPFNQKISVGTSHENDKGYDLEPDWQVLKKLEMQALTYLPLLKEATQKTCRVGIRAY 301
Query: 303 TSDSFPPFFGPVPCMEGAYAASGLGSTGLTVGPLIGYELCQLILNKENQLNLEDYDITKYV 362
TSD+SPF+G V ++ Y ASGLGS+GLTVGPLIGYEL QL+L E L DY Y+
40 Sbjct: 302 TSDYSFPYQVSGLKNLYTASGLGSSGLTVGPLIGYELAQLLLGHEGLLTPSDYSPEPYL 361

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
vaccines or diagnostics.

A related GBS gene <SEQ ID 8679> and protein <SEQ ID 8680> were also identified. Analysis of this
protein sequence reveals the following:

45 Lipop Possible site: -1 Crend: 2
McG: Discrim Score: 4.44
GvH: Signal Score (-7.5): 0.81
Possible site: 41
50 >>> Seems to have a cleavable N-term signal seq.
ALOM program count: 0 value: 7.32 threshold: 0.0
PERIPHERAL Likelihood = 7.32 153
modified ALOM score: -1.96
*** Reasoning Step: 3
55 ----- Final Results -----
bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
60 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

Bacillus subtilis

SEQ ID 8680 (GBS290) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 57 (lane 6; MW 22kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 77 (lane 4; MW 47kDa).

GBS290-GST was purified as shown in Figure 226, lane 9.

40 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 875

A DNA sequence (GBSx0927) was identified in *S.agalactiae* <SEQ ID 2657> which encodes the amino acid sequence <SEQ ID 2658>. Analysis of this protein sequence reveals the following:

```

45      Possible site: 20
      >>> Seems to have no N-terminal signal sequence
          INTEGRAL      Likelihood = -2.18      Transmembrane      38 - 54 ( 36 - 54)

      ----- Final Results -----
50      bacterial membrane --- Certainty=0.1871(Affirmative) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

55 >GP:AAD19913 GB:AF105113 glucose-1-phosphate thymidyl transferase
[*Streptococcus pneumoniae*]
Identities = 262/289 (90%), Positives = 276/289 (94%)

-967-

Query: 1 MKGIILAGSGSTRLYPLTRAASKQLMPIYDKPMIYYPLSVLMLAGIKEILIISTPQDLPR 60
 MKGIILAGSGSTRLYPLTRAASKQLMP+YDKPMIYYPLS LMLAGIK+LIISTPQDLPR
 Sbjct: 1 MKGIILAGSGSTRLYPLTRAASKQLMPVYDKPMIYYPLSTLMLAGIKDILIISTPQDLPR 60

5 Query: 61 FEDMLGDGSELGISLSYAEQSPDGLAQAFIIGEDFIGDDHVALVLGDNIYHGPGLSAML 120
 F+D+L DGSE GI LSYAEQSPDGLAQAF+IGE+FIGDD VAL+LGDNIYHGPGLS ML
 Sbjct: 61 FKDLLLDGSEFGIKLSYAEQSPDGLAQAFIIGEEFIGDDSVAILLGDNIYHGPGLSTML 120

10 Query: 121 QRAASKESGATVFGYQVKDPERFGVVEFDITDMNAISIEEKPAQPKSNYAVTGLYFYDNDV 180
 Q+AA KE GATVFGYQVKDPERFGVVEFDITDMNAISIEEK P+SNYAVTGLYFYDNDV
 Sbjct: 121 QKAAKKEKGATVFGYQVKDPERFGVVEFDITDMNAISIEEKPEYPRSNYAVTGLYFYDNDV 180

15 Query: 181 VEIAKNIKPSRGELEITDVNKAYLDRGDLSELVLMGRGFAWLDTGTTHESLLEAAQYIETV 240
 VEIAK IKPS RGELEITDVNKAYL+RGDLSVELMGRGFAWLDTGTTHESLLEA+QYIETV
 Sbjct: 181 VEIAKQIKPSRGELEITDVNKAYLNRGDLSELVLMGRGFAWLDTGTTHESLLEASQYIETV 240

20 Query: 241 QRMQNVQVANLEEIAYRMGYITREQVLELAQPLKKNEYGQYLLRLIGEA 289
 QRMQNVQVANLEEI+YRMGYI+RE VLELAQPLKKNEYG+YLLRLIGEA
 Sbjct: 241 QRMQNVQVANLEEISYRMGYISREDVLELAQPLKKNEYGRYLLRLIGEA 289

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2659> which encodes the amino acid sequence <SEQ ID 2660>. Analysis of this protein sequence reveals the following:

Possible site: 20
 >>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1585 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

30 RGD motif: 207-209

The protein has homology with the following sequences in the databases:

>GP:AAC69538 GB:AF057294 Cps23f0 [Streptococcus pneumoniae]
 Identities = 263/289 (91%), Positives = 276/289 (95%)

35 Query: 1 MKGIILAGSGSTRLYPLTRAASKQLMPIYDKPMIYYPLSTLMLAGIKDVIISTPQDLPR 60
 MKGIILAGSGSTRLYPLTRAASKQLMP+YDKPMIYYPLSTLMLAGI+D+LIISTPQDLPR
 Sbjct: 1 MKGIILAGSGSTRLYPLTRAASKQLMPVYDKPMIYYPLSTLMLAGIRDILIISTPQDLPR 60

40 Query: 61 FEELLGDGSEFGISLSYKEQSPDGLAQAFIIGEEFIGDDRVALILGDNIYHGNLTKML 120
 F+ELL DGSEFGI LSY EQSPDGLAQAFIIGEEFIGDD VALILGDNIYHG GL+ ML
 Sbjct: 61 FKELLQDGSEFGIKLSYAEQSPDGLAQAFIIGEEFIGDDSVAILLGDNIYHGPGLSTML 120

45 Query: 121 QKAAAKEKGATVFGYQVKDPERFGVVEFDENMNAISIEEKPEVPKSHFAVTGLYFYDNDV 180
 QKAA KEGATVFGY VKDPERFGVVEFDENMNAISIEEKPE P+S++AVTGLYFYDNDV
 Sbjct: 121 QKAAKKEKGATVFGYHVKDPERFGVVEFDENMNAISIEEKPEYPRSNYAVTGLYFYDNDV 180

50 Query: 181 VEIAKNIKPSARGELEITDVNKAYLBERGDLSELVLMGRGFAWLDTGTTHESLLEAAQYIETV 240
 VEIAK+IKPS RGELEITDVNKAYL+RGDLSVELMGRGFAWLDTGTTHESLLEA+QYIETV
 Sbjct: 181 VEIAKSIKPSRGELEITDVNKAYLDRGDLSELVLMGRGFAWLDTGTTHESLLEASQYIETV 240

55 Query: 241 QRLQNAQVANLEEIAYRMGYISKEDVHKLAQSLKKNEYGQYLLRLIGEA 289
 QR+QN QVANLEEIAYRMGYIS+EDV LAQSLKKNEYGQYLLRLIGEA
 Sbjct: 241 QRMQNVQVANLEEIAYRMGYISREDVLALAQSLKNEYGQYLLRLIGEA 289

An alignment of the GAS and GBS proteins is shown below.

Identities = 257/289 (88%), Positives = 274/289 (93%)

60 Query: 1 MKGIILAGSGSTRLYPLTRAASKQLMPIYDKPMIYYPLSVLMLAGIKEILIISTPQDLPR 60
 MKGIILAGSGSTRLYPLTRAASKQLMPIYDKPMIYYPLS LMLAGIK++LIISTPQDLPR
 Sbjct: 1 MKGIILAGSGSTRLYPLTRAASKQLMPIYDKPMIYYPLSTLMLAGIKDVIISTPQDLPR 60

Query: 61 FEDMLGDGSELGISLSYAEQSPDGLAQAFIIGEDFIGDDHVALVLGDNIYHGPGLSAML 120

-968-

FE++LGDGSE GISLSY EQSPDGLAQAFIIGE+FIGDD VAL+LGDNIYHG GL+ ML
 Sbjct: 61 FEELLGDGSEFGISLSYKEQSPDGLAQAFIIGEEFIGDDRVALILGDNIYHGNGLT'KML 120

Query: 121 QRAASKESGATVFGYQVKDPERFGVVEFD+MNAISIEEKPAQPKSNYAVTGLYFYDNDV 180
 Q+AA+KE GATVFGYQVKDPERFGVVEFD +MNAISIEEK PK++AVTGLYFYDNDV
 Sbjct: 121 QKAAAKEKGATVFGYQVKDPERFGVVEFDENMNAISIEEKPEVVPKSHFAVTGLYFYDNDV 180

Query: 181 VEIAKNIKPSRGELEITDVNKAYLDRGDLVELMGRGFAWLDTGTHESLLEAAQYIETV 240
 VEIAKNIKPS RGELEITDVNKAYL+RGDLVELMGRGFAWLDTGTHESLLEAAQYIETV
 Sbjct: 181 VEIAKNIKPSARGELEITDVNKAYLERGDLVELMGRGFAWLDTGTHESLLEAAQYIETV 240

Query: 241 QRMQNQVANLEEIAYRMGYITREQVLELAQPLKKNEYGQYLLRLIGEA 289
 QR+QN QVANLEEIAYRMGYI++E V +LAQ LKKNEYGQYLLRLIGEA
 Sbjct: 241 QRLQNAQVANLEEIAYRMGYISKEDVHKLQSLKKNEYGQYLLRLIGEA 289

There is also homology to SEQ ID 858.

SEQ ID 2658 (GBS296) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 43 (lane 5; MW 35.4kDa).

GBS296-His was purified as shown in Figure 203, lane 7.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 876

A DNA sequence (GBSx0929) was identified in *S.agalactiae* <SEQ ID 2661> which encodes the amino acid sequence <SEQ ID 2662>. Analysis of this protein sequence reveals the following:

Possible site: 18
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2635(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 877

A DNA sequence (GBSx0930) was identified in *S.agalactiae* <SEQ ID 2663> which encodes the amino acid sequence <SEQ ID 2664>. This protein is predicted to be unnamed protein product. Analysis of this protein sequence reveals the following:

Possible site: 56
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1868(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2665> which encodes the amino acid sequence <SEQ ID 2666>. Analysis of this protein sequence reveals the following:

-969-

Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.2818(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

RGD motif: 29-31

10

The protein has homology with the following sequences in the databases:

>GP:AAC69539 GB:AF057294 Cps23fP [Streptococcus pneumoniae]
Identities = 168/197 (85%), Positives = 183/197 (92%)

15 Query: 1 MTETFFDKPLACREIKEIPGLLEFDIPVRGDNRGWFKENFQKEKMLPIGFPERFFEEGKL 60
 MT+ FF K LA R+++ IPG+LEFDIPV GDNRGWFKENFQKEKMLP+GFPE FF EGGLE
 Sbjct: 1 MTDNFFGKTLAARKVEAIPGMLLEFDIPVHGDNRGWFKENFQKEKMLPLGFPEFFAEGKL 60

20 Query: 61 QNNVSFSRQHVLRLGLHAEPWDKYISVADDGKVLGAWVDLREGETFGNVYQTVIDASKGMF 120
 QNNVSFSR++VLRGLHAEPWDKYISVAD GKVLG+WVDLREGETFGN YQTVIDASKG+F
 Sbjct: 61 QNNVSFSRKNVLRGLHAEPWDKYISVADGGKVLGSWVDLREGETFGNTYQTVIDASKGIF 120

25 Query: 121 VPRGVANGFQVLSSETVSYSYLVNDYWALDLKPKYAFVNYADPSLGITWENLAAAEVSEAD 180
 VPRGVANGFQVLS+TVSYSYLVNDYWAL+LKPKYAFVNYADPSLGI WEN+A AEVSEAD
 Sbjct: 121 VPRGVANGFQVLSDTVSYSLVNDYWALELKPKYAFVNYADPSLGIWENLAAAEVSEAD 180

 Query: 181 KNHPLLSDVKPLPKDL 197
 K+HPLL DVKPLK +DL
 Sbjct: 181 KHHPLLKDVKPLKEDL 197

30

An alignment of the GAS and GBS proteins is shown below.

Identities = 157/197 (79%), Positives = 180/197 (90%)

35 Query: 1 MTEQFFDKELTCRPIEAIPGLLEFDIPVRGDNRGWFKENFQKEKMLPIGFPEFFFEADKL 60
 MTE FFDK L CR I+ IPGLLEFDIPVRGDNRGWFKENFQKEKM+P+GFPE FFE KL
 Sbjct: 1 MTETFFDKPLACREIKEIPGLLEFDIPVRGDNRGWFKENFQKEKMLPIGFPERFFEEGKL 60

40 Query: 61 QNNISFNKNTLRGLHAEPWDKYISVADGKVLGAWVDLREGDSFGNVYQTVIDASKGIF 120
 QNN+SF++++ LRGLHAEPWDKY+S+AD+G+V+G WVDLREG++FGNVYQT+IDASKG+F
 Sbjct: 61 QNNVSFSRQHVLRLGLHAEPWDKYISVADDGKVLGAWVDLREGETFGNVYQTVIDASKGMF 120

45 Query: 121 VPRGVANGFQVLSKAAAYTYLVNDYWALELKPKYAFVNYADPNLGIQWENLEAAEVSEAD 180
 VPRGVANGFQVLS+ +Y+YLVNDYWAL+LKPKYAFVNYADP+LGI WENL AEVSEAD
 Sbjct: 121 VPRGVANGFQVLSSETVSYSYLVNDYWALDLKPKYAFVNYADPSLGITWENLAAAEVSEAD 180

 Query: 181 KNHPLLKDVKPLKEDL 197
 KNHPLL DVKPLK +DL
 Sbjct: 181 KNHPLLSDVKPLPKDL 197

50 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 878

A DNA sequence (GBSx0931) was identified in *S.galactiae* <SEQ ID 2667> which encodes the amino acid sequence <SEQ ID 2668>. Analysis of this protein sequence reveals the following:

55 Possible site: 14
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

60 bacterial cytoplasm --- Certainty=0.3019(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-970-

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 879

A DNA sequence (GBSx0932) was identified in *S.agalactiae* <SEQ ID 2669> which encodes the amino acid sequence <SEQ ID 2670>. Analysis of this protein sequence reveals the following:

```

Possible site: 37
10  >>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside  --- Certainty=0.0000(Not Clear) < succ>
15  bacterial cytoplasm  --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 880

A DNA sequence (GBSx0933) was identified in *S.agalactiae* <SEQ ID 2671> which encodes the amino acid sequence <SEQ ID 2672>. Analysis of this protein sequence reveals the following:

```

Possible site: 38
25  >>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.0957(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
30  bacterial outside  --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9367> which encodes amino acid sequence <SEQ ID 9368> was also identified.

The protein is similar to the dTDP-glucose-4,6-dehydratase from *S.mutans*:

```

35  >GP:BAA11249 GB:D78182 dTDP-glucose-4,6-dehydratase [Streptococcus mutans]
      Identities = 290/310 (93%), Positives = 304/310 (97%)

Query: 1  MTYAGN RANIEAILGDRVELVVGDIADAELVDKLAAKADAIVHYAESHNDNSLNDPSPF 60
      +TYAGN AN+E ILGDRVELVVGDIAD+ELVDKLAAKADAIVHYAESHNDNSL DPSPF
40  Sbjct: 39 LTYAGN HANLEEILGDRVELVVGDIADSELVDKLAAKADAIVHYAESHNDNSLKDPSPF 98

Query: 61 IHTNFIGTYTLLEAARKYDIRFHHVSTDEVYGDPLREDLPGNGEGPGEKFTAETKYNPS 120
      I+TNF+CTYTLLEAARKYDIRFHHVSTDEVYGDPLREDLPG+GEGPGEKFTAETKYNPS
Sbjct: 99 IYTNFVGTYTLLEAARKYDIRFHHVSTDEVYGDPLREDLPGHGEKFTAETKYNPS 158

45  Query: 121 SPYSSTKAASDLIVKAWVRSFGVKATISNCSNNGPYQHIEKFIPRQITNLAGIKPKLY 180
      SPYSSTKAASDLIVKAWVRSFGVKATISNCSNNGPYQHIEKFIPRQITNLAGIKPKLY
Sbjct: 159 SPYSSTKAASDLIVKAWVRSFGVKATISNCSNNGPYQHIEKFIPRQITNLAGIKPKLY 218

50  Query: 181 GEGKNVRDWIHTNDHSTGVWAILTKGRIGETYTLIGADGEKNNKEVLELILEKMGQPKDAY 240

```

-971-

GEGKNVRDWIHTNDHSTGVWAILTKGRIGETYLIGADGEKNNKEVLELILEKM QPKDAY
 Sbjct: 219 GEGKNVRDWIHTNDHSTGVWAILTKGRIGETYLIGADGEKNNKEVLELILEKMSQPKDAY 278
 Query: 241 DHVTDRAHDLRYAIDSTKLREELGWEPQFTNFSEGLEETINWYTENQDWWKAEKEAVEA 300
 5 DHVTDRAHDLRYAIDSTKLREELGW+PQFTNF EGLE+TI WYTE++DWWKAEKEAVEA
 Sbjct: 279 DHVTDRAHDLRYAIDSTKLREELGWKPQFTNFEEGLEDTIKWYTEHEDWWKAEKEAVEA 338
 Query: 301 NYAKTQEVIN 310
 NYAKTQ+++N
 10 Sbjct: 339 NYAKTQKILN 348

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2673> which encodes the amino acid sequence <SEQ ID 2674>. Analysis of this protein sequence reveals the following:

Possible site: 40
 15 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1150(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 20 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 300/309 (97%), Positives = 303/309 (97%)

25 Query: 1 MTYAGNRANIEAILGDRVELVVGDIADAELVDKLAAKADAIVHYAESHNDNSLNDPSPF 60
 +TYAGNRANIEAILGDRVELVVGDIADAELVDKLAAK DAIVHYAESHNDNSL DPSPF
 Sbjct: 37 LTYAGNRANIEAILGDRVELVVGDIADAELVDKLAAKTDAIVHYAESHNDNSLEDPSPF 96
 Query: 61 IHTNFIGTYTLLEAARKYDIRFHHVSTDEVYGDLPREDLPNGEGPGEKFTAETKYNPS 120
 30 IHTNFIGTYTLLEAARKYDIRFHHVSTDEVYGDLPREDLPG GEGPGEKFTAETKYNPS
 Sbjct: 97 IHTNFIGTYTLLEAARKYDIRFHHVSTDEVYGDLPREDLPQGEGPGEKFTAETKYNPS 156
 Query: 121 SPYSSTKAASDLIVKAWVRSFGVKATISNCSNNYGPYQHIEKFIPRQITNLAGIKPKLY 180
 SPYSSTKAASDLIVKAWVRSFGVKATISNCSNNYGPYQHIEKFIPRQITNLAGIKPKLY
 35 Sbjct: 157 SPYSSTKAASDLIVKAWVRSFGVKATISNCSNNYGPYQHIEKFIPRQITNLAGIKPKLY 216
 Query: 181 GEGKNVRDWIHTNDHSTGVWAILTKGRIGETYLIGADGEKNNKEVLELILEKMGQPKDAY 240
 GEGKNVRDWIHTNDHSTGVWAILTKGRIGETYLIGADGEKNNKEVLELILEKMGQPKDAY
 40 Sbjct: 217 GEGKNVRDWIHTNDHSTGVWAILTKGRIGETYLIGADGEKNNKEVLELILEKMGQPKDAY 276
 Query: 241 DHVTDRAHDLRYAIDSTKLREELGWEPQFTNFSEGLEETINWYTENQDWWKAEKEAVEA 300
 DHVTDRAHDLRYAIDSTKLREELGWEPQFTNFSEGLEETI WYTEN+ WWKAEK+AVEA
 Sbjct: 277 DHVTDRAHDLRYAIDSTKLREELGWEPQFTNFSEGLEETIKWYTENETWWKAEKDAVEA 336
 45 Query: 301 NYAKTQEVI 309
 YAKTQEVI
 Sbjct: 337 KYAKTQEVI 345

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 50 vaccines or diagnostics.

Example 881

A DNA sequence (GBSx0935) was identified in *S.agalactiae* <SEQ ID 2675> which encodes the amino acid sequence <SEQ ID 2676>. Analysis of this protein sequence reveals the following:

Possible site: 36
 55 >>> Seems to have a cleavable N-term signal seq.
 ----- Final Results -----
 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 60 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 882

A DNA sequence (GBSx0936) was identified in *S.agalactiae* <SEQ ID 2677> which encodes the amino acid sequence <SEQ ID 2678>. Analysis of this protein sequence reveals the following:

```

Possible site: 35
>>> Seems to have an uncleavable N-term signal seq
    INTEGRAL    Likelihood =-15.55    Transmembrane    13 - 29 ( 3 - 40)

----- Final Results -----
    bacterial membrane --- Certainty=0.7220(Affirmative) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 883

A DNA sequence (GBSx0937) was identified in *S.agalactiae* <SEQ ID 2679> which encodes the amino acid sequence <SEQ ID 2680>. Analysis of this protein sequence reveals the following:

```

Possible site: 15
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
    bacterial cytoplasm --- Certainty=0.2882(Affirmative) < succ>
    bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 884

A DNA sequence (GBSx0938) was identified in *S.agalactiae* <SEQ ID 2681> which encodes the amino acid sequence <SEQ ID 2682>. This protein is predicted to be hyaluronate lyase. Analysis of this protein sequence reveals the following:

```

Possible site: 30
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
    bacterial outside --- Certainty=0.3000(Affirmative) < succ>

```

-973-

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2683> which encodes the amino acid
 5 sequence <SEQ ID 2684>. Analysis of this protein sequence reveals the following:

Possible site: 46
 >>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

10 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related sequence was also identified in GAS <SEQ ID 9099> which encodes the amino acid sequence
 15 <SEQ ID 9100>. Analysis of this protein sequence reveals the following:

Possible cleavage site: 23
 >>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

20 bacterial outside --- Certainty= 0.300(Affirmative) < succ>
 bacterial membrane --- Certainty= 0.000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty= 0.000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

25 Identities = 359/771 (46%), Positives = 492/771 (63%), Gaps = 50/771 (6%)

Query: 307 PNAT--GSTTVKISDKSGKIIKEVPLSVTASTEDNFTKLLDKWNDVTIGNHVYDTNDSNM 364
 PN T + T+ +D K+++ +D +T+LLD+WN + GN YD + +M
 30 Sbjct: 65 PNNTYFQTQTLTTTDDSEKKVVQP-----QQKDYVTELLDQWNSIIAGNDAYDKTNPDM 117

Query: 365 QKLNQKLDETNAKNIEAIKL-----DSNRTFLWKDLNLSNSAQLTATYRRLDLAKQIT 419
 + K E +A+NI IK NRT+LW+ + + SA +T TYR +E +AKQIT
 35 Sbjct: 118 VTFHNKA-EKDAQNI--IKSYQGPDPHENRTYLWEHAKDYSASANTTKTYRNIKIAKQIT 174

Query: 420 NPHSTIYKNEKAIRTVKESLAWLHQNFYVNVNKDI-----EGSANWWDPEIGVPRISITGT 473
 NP S Y++ KAI VK+ +A++++ YN++++ E NWW +EIG PR+I T
 40 Sbjct: 175 NPESCYQDSKAIAIVKDGMAFMYEHAYNLDRENHQTGKENKENWVVEIGTFRAINNT 234

Query: 474 LALMNYFTDAEIKTYTDPIEHFVPDAGFFRKTLVN--PFKALGGLVDMGRVKIIEGLL 531
 L+LMY YFT EI YT PIE FVPD FR N PF+A GNL+DMGRVK+I G+L
 45 Sbjct: 235 LSLMPYPTQEEILKYTAPIEKFPVDPTRFRVRAANFSPFEANSGLIDMGRVKLISGIL 294

Query: 532 RKDNTIIEKTSLSLKNLFTTATKAEGFYADGSYIDHT-----NVAYTGAYGNVL 580
 RKD+ I T +++ +FT + GFY DGS IDH +AYTGAYGNVL
 50 Sbjct: 295 RKDDLEISDTIKAIEKVFTLVDEGNGFYQDGLIDHVVTNAQSPPLYKKGIAYTGAYGNVL 354

Query: 581 IDGLTQLLPPIIQETDYKISNQELDMVYKWINQSFPLPLIVKGELMDMSRGRSISREAASSH 640
 IDGL+QL+PIIQ+T I ++ +Y WIN SF P+IV+GE+MDM+RGRSISR A SH
 55 Sbjct: 355 IDGLSQLIPIIQKTKSPIKADKMATIYHWINHSFFPIIVRGEMDMTRGRSISRFAQSH 414

Query: 641 AAAVEVLRGFLRLANMSNEERNLDLKSTIKTIITS-NKFYNVFNLSYSDIANMNKLLN 699
 A +E LR LR+A+MS E L LK+ IKT++T N FYNV++NLK+Y DI M +LL+
 60 Sbjct: 415 VAGIEALRAILRIADMSEEPHRLALKTRIKTLVTQGNAFYVYDNLKTYHDIKLMKELLS 474

Query: 700 DSTVATKPLKSNLSTFNSMDRLAYYNAEKDFGFALSLSKRTLNYEGMNDENTRGWYITGD 759
 D++V + L S +++FNSMD+LA YN + DF F LS+ S RT NYE MN+EN GW+T D
 65 Sbjct: 475 DTSVPVQKLDYSYVASFNSMDKLALYNNKHDFAFGLSMFSNRTQNYEAMNNENLHGWFSTD 534

Query: 760 GMFYLYNSDQSHYSNHFWPTVNPYKMGATTEKDAKREDTTTKDFMSKHSKDAKEKTGQVTG 819
 GMFYLYN+D HYS ++W TVNPY++ GTTE + K + T + + K ++ G +TG
 70 Sbjct: 535 GMFYLYNNDLGHYSENWATVNPYRLPGTTETEOKPLEGTPE----NIKINYQQVG-MTG 589

Query: 820 ASD--FVGSVKLNDHFALAAMDFTNWDRTLTQAQKGWVILNDKIVFLGSNIKNTNGIGNVS 877

-974-

SD FV S KIN+ ALAAM FTNW+++LT KGW IL +KI+F+GSNIKN +
 Sbjct: 590 LSDDAFVASKKLNNTSALAAMFTFNWKS LTLNKGWIFLGNKIIFVGSNIKNQSS-HKAY 648

Query: 878 TTIDQRKDDSKTPYTTYVNGKTVDLKQASSQQTDTKSVFLESKEPGRNIGYIFFKNSTI 937
 TTI+QRK++ K PY +YVN + VDL FT+TKS+FLES +P +NIGY FFK +T+
 Sbjct: 649 TTIEQRKENQKYPYCSYVNNQPVLDLN-QLVDFITNTKSIFLESDDPAQNIGYFFKPTTL 707

Query: 938 DIERKEQTGTWNSINRTSKNTSI--VSNPFITISQKHDNKGDSYDMMVPNIDRTSFDK 994
 I + QTG W +I K+ VSN FITI Q H GD Y YMM+PN+ R F+
 Sbjct: 708 SISKALQTGWQNIKADDDKSPEAIKEVSNTFITIMQNHTQDGRYAYMMLPNMTRQEFET 767

Query: 995 LANSKEVELLENSKQQVIYDKNSQTWAVIKHDNQESLINNQFKMNKAGLY 1045
 + +++LLEN+ K +YD +SQ VI + + ++ +N ++ G Y
 Sbjct: 768 YISKLDIDLLENNDKLAAYVDHDSQQMHVIHYGKKATMFSNH-NLSHQGFY 817

SEQ ID 2682 (GBS89) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 6 (lane 3; MW 118kDa).

The His-fusion protein was purified as shown in Figure 190, lane 4.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 885

A DNA sequence (GBSx0939) was identified in *S.galactiae* <SEQ ID 2685> which encodes the amino acid sequence <SEQ ID 2686>. This protein is predicted to be mutator mutt protein. Analysis of this protein sequence reveals the following:

Possible site: 42
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3781(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAA11250 GB:D78182 MutX [Streptococcus mutans]
 Identities = 132/160 (82%), Positives = 146/160 (90%), Gaps = 1/160 (0%)

Query: 1 MTKLATICYIDNGKELLLHRNKKENDVHEGKWISVGGKLEAGETPDECAKREILLEETHL 60
 M KLATICYIDNG+ELL+HRNKK NDVHEGKWISVGGKLE GE+PDECA+REI EETHL
 Sbjct: 1 MIKLATICYIDNGRELLLMHRNKKENDVHEGKWISVGGKLEKGESPDCEARREIFEETHL 60

Query: 61 TVKKMDFKGVITFPFTPGHDWYTYVFKVTDYEGELISDDDESREGTLEWVPYDQVLSKPT 120
 VK+MDFKG+ITFP+FTPGHDWYTYVFKV D+EG LISD +SREGTLEWVPY+QVL+KPT
 Sbjct: 61 IVKQMDFKGIITFPDFTPGHDWYTYVFKVRDFEGR LISDKDSREGTLEWVPYNQVLT KPT 120

Query: 121 WQGDYEIFKWILEDVPPFSAKFVYDEHQNLIEKTVNPFYEK 160
 W+GDYEIFKWILED PFFSAKFVY E Q L++K V FYEK
 Sbjct: 121 WEGDYEIFKWILEDAPFFSAKFVYQE-QKLVDKHFVIFYEK 159

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2687> which encodes the amino acid sequence <SEQ ID 2688>. Analysis of this protein sequence reveals the following:

Possible site: 42
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3399(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

-975-

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 131/158 (82%), Positives = 146/158 (91%)

5 Query: 1 MTKLATICYIDNGKELLLLHRNKKENDVHEGKWISVGGKLEAGETPDECAKREILEETHL 60
 MT+LATICYIDNG LLLLHRNKKENDVH+GKWISVGGKLEAGETPDECA+REILEETHL
 Sbjct: 1 MTQLATICYIDNGDSL LLLLHRNKKENDVHKGWISVGGKLEAGETPDECARREILEETHL 60

10 Query: 61 TVKKMDFKGVITFPFETPGHDWYTYVFKVTDYEGELISDDESREGTLEWVPYDQVLSKPT 120
 TV +M FKG+ITFPFETPGHDWYTYVFKVT +EG+LISD+ESREGTLEWVPYDQVL KPT
 Sbjct: 61 TVTEMAFKGIITFPFETPGHDWYTYVFKVTGFEGDLISDEESREGTLEWVPYDQVLEKPT 120

15 Query: 121 WQGDYEIFKWILEDVPPFSAKFVYDEHQNLIEKTVNFY 158
 W+GDY+IFKWILED FPSAKF YD++ L++K+V FY
 Sbjct: 121 WEGDYDIFKWILED RSPFSAKFYTDQNNQLMDKSVTFY 158

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

20 Example 886

A DNA sequence (GBSx0940) was identified in *S.agalactiae* <SEQ ID 2689> which encodes the amino acid sequence <SEQ ID 2690>. This protein is predicted to be MutT/nudix family protein. Analysis of this protein sequence reveals the following:

25 Possible site: 28
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1901 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

35 >GP:AAF11817 GB:AE002059 MutT/nudix family protein [Deinococcus radiodurans]
 Identities = 40/135 (29%), Positives = 62/135 (45%), Gaps = 3/135 (2%)

 Query: 22 FGVRVSALIIENQKLLLIYAPHLDKYY-LPGGALQVGEDSNKAVAREVLEEIGLHSQVGD 80
 F R + + + + +LL + ++ LPGGA+Q GE S A RE EE GL + V
 Sbjct: 33 FQTRATLICVDNRLLTCWDERFPDFFALPGGAVQTGESSAAAAQREWHETGLRADVTR 92

40 Query: 81 LAYIIENQFNKRHHYHSVEFLYFVNLLGQAPESIKEGTHKRHFVWLPIKELTKIDCNPN 140
 A +E F+ + H F + V L G+ P + + + H F WL + L P
 Sbjct: 93 CA-TLERFFHWEGRRERHEFGFFRVELTGELPATVLDNPHV-PFRWLAVDALDDHTLYPR 150

45 Query: 141 FLAQDLIEWPGHVH 155
 + Q L G + H
 Sbjct: 151 CVFQLRLRLPAGEIGH 165

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2691> which encodes the amino acid sequence <SEQ ID 2692>. Analysis of this protein sequence reveals the following:

50 Possible site: 55
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3832 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

-976-

Identities = 33/80 (41%), Positives = 50/80 (62%), Gaps = 1/80 (1%)

Query: 29 LIIENQKLLLIYAPHLDKYYLPGGALQVGEDSNKAVAREVLEEIGLHSQVGDLAYIIENQ 88

LI+ N K L D+YY GG VGE +++ V RE LEE+G+ ++V LA+++EN

Sbjct: 1 LIVRNGKNFLTRDAD-DQYYTIGGTSLVGEKTHETVLRETLEEVGIRAKVNQLAFMVENH 59

Query: 89 FNIKRHHYHSVEFLYFVNLL 108

F+I +H++EF Y V+ L

Sbjct: 60 FDIDDVFWHNIEFHYLVSP 79

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 887

A DNA sequence (GBSx0941) was identified in *S.agalactiae* <SEQ ID 2693> which encodes the amino acid sequence <SEQ ID 2694>. This protein is predicted to be unnamed protein product. Analysis of this protein sequence reveals the following:

Possible site: 26

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -12.95 Transmembrane 24 - 40 (17 - 48)

INTEGRAL Likelihood = -11.09 Transmembrane 88 - 104 (82 - 112)

INTEGRAL Likelihood = -9.39 Transmembrane 294 - 310 (276 - 315)

INTEGRAL Likelihood = -8.07 Transmembrane 242 - 258 (236 - 262)

INTEGRAL Likelihood = -7.86 Transmembrane 50 - 66 (43 - 74)

INTEGRAL Likelihood = -3.13 Transmembrane 337 - 353 (332 - 355)

INTEGRAL Likelihood = -2.23 Transmembrane 185 - 201 (182 - 202)

INTEGRAL Likelihood = -1.38 Transmembrane 269 - 285 (267 - 285)

----- Final Results -----

bacterial membrane --- Certainty=0.6180(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2695> which encodes the amino acid sequence <SEQ ID 2696>. Analysis of this protein sequence reveals the following:

Possible site: 26

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -9.71 Transmembrane 88 - 104 (85 - 112)

INTEGRAL Likelihood = -9.29 Transmembrane 24 - 40 (21 - 72)

INTEGRAL Likelihood = -8.92 Transmembrane 47 - 63 (41 - 72)

INTEGRAL Likelihood = -7.59 Transmembrane 243 - 259 (237 - 266)

INTEGRAL Likelihood = -6.10 Transmembrane 181 - 197 (178 - 203)

INTEGRAL Likelihood = -5.47 Transmembrane 278 - 294 (273 - 310)

INTEGRAL Likelihood = -3.88 Transmembrane 338 - 354 (331 - 368)

INTEGRAL Likelihood = -1.59 Transmembrane 297 - 313 (297 - 314)

----- Final Results -----

bacterial membrane --- Certainty=0.4885(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAD00285 GB:U78604 putative membrane protein [Streptococcus mutans]

Identities = 244/382 (63%), Positives = 310/382 (80%), Gaps = 3/382 (0%)

Query: 12 SLFYKWFLLNQATMALVITLLAFLTIFVFTKISFLFMPVISFFAVIMLPLVISTILYYLT 71

S F+KWFL+N+ I++ LL FL I VFTKIS +F P++SF AVIMLPLVIS +LYYL

Sbjct: 17 SWFFKWFLLDNKTVTVLLVLLVFLDILVFTKISSIFKPLLSFLAVIMLPLVISALLYLL 76

Query: 72 KPLVDLINHLGPNRTTSIFIVFGLITLLFVWALSGFVPMVQTQLTSFIEDLPKYVGKVNE 131

-977-

KP+VD I G +R +I IVF +I L VW I+ F PM+ QLTSFI+ LP YV V+
 Sbjct: 77 KPIVDFIEIRGTSRVMAITIVFVIIAGLLVWGIANFFPMLNEQLTSFIKYLPYSYVRSVDA 136
 Query: 132 EANKLLENEWLVSYKPOLQDMLTHTSQKALDYAQSFSKNAIDWAGNFAGAIARITVAIII 191
 + +KLL N+ L S++PQ+++ +T+ SQKA+DYA+ FSK A+ WAGNFA IAR+TVAIIII
 Sbjct: 137 QVSKLLRNDLLASFRPQIENAVTNFSQKAVDYAEPFSKGAVTWAGNFASLIARVTVAIII 196
 Query: 192 SPFILFYFLRDSSHMKNGLVNVLPKLRVPMVRVLGDINKQLSGYVQGQVTVAIIVGFMF 251
 SPFI+FY LRDSS MK V+ LP K+R P+ R+LGD+N+QL+GYVQ TVAI+VGFMF
 Sbjct: 197 SPFIVFYLLRDSSMKKEAFVSYPKMRQPIHRIILGDVNRQLAGYVQSRSTVAIIVGFMF 256
 Query: 252 SIMFSLVGLKYAITFGIAGFLNMIPYLGSLAMIPVIMAMVQGPFMLVKVLVIFMIEQ 311
 SIMF+++GL+YA+TFGIAGFLNMIPYLGSLA IPV I+A+V+GP +VKV ++F++EQ
 Sbjct: 257 SIMFTIIGLRYAVTFGIAGFLNMIPYLGSLATIPVFILALVEGPKVVKVALVFIVEQ 316
 Query: 312 TIEGRFVAPLVGNKLSIHPITIMFLLTAGSMFGVWGVFLVIPIYASVKVVIKELFDWY 371
 TIEGRFV+PLVLG+KLSIHPITIMF+LLTAGSMFGVWGVFL IP+YAS+KVV+KE+F+WY
 Sbjct: 317 TIEGRFVSPLVLGSKLSIHPITIMFILLTAGSMFGVWGVFLGIPVYASIKVVVKEIFEWY 376
 Query: 372 KKVSGLYDEEVLVIEEVKDHVK 393
 K +SGLY++E E++K VK
 Sbjct: 377 KPISGLYEKEE---EDIKKDVK 395

An alignment of the GAS and GBS proteins is shown below.

Identities = 243/389 (62%), Positives = 306/389 (78%), Gaps = 2/389 (0%)
 Query: 6 EKEFKNSLFKKWILNNQAVIALMITFLVFLTIFITKISFMFKPVDFLAVLILPLVISG 65
 EK +SLF+KW LNNQA +AL+IT L FLTIF+FTKISF+F PV F AV++LPLVIS
 Sbjct: 6 EKSRTDSLFFYKWLNNQATMALVITLLAFLTIFVFTKISFLFMPVISFFAVIMLPLVIST 65
 Query: 66 LLYYLLKPMVTLEKRGIKRVTAILSVTIIILLIWMSSFIPMNSQLRHFMEDLPSY 125
 +LYYL KP+V + G R T+I VF +I LL +WA+S F+PM+ QL F+EDLP Y
 Sbjct: 66 ILVYLTKPLVDLINHLGNRTTSIFIVFGLITILFVWALSFGVPMVQTQLTSFIEDLPKY 125
 Query: 126 VNKVQMETSSFIDHNPWLKSKYGEISSMLSNISSQAVSYAEKFSKNILDWAGNLA TVAR 185
 V KV E + ++ N WL SYK ++ ML++ S +A+ YA+ FSKN +DWAGN A +AR
 Sbjct: 126 VGKVNNEEANKLLE-NEWLVSYKPOLQDMLTHTSQKALDYAQSFSKNAIDWAGNFAGAIAR 184
 Query: 186 VTVATIMAPFILFYLLRDSRNMGFLMVLPTKLRQPTDRILREMNSQMSGYVQGQIIVA 245
 +TVA I++PFILFY LRDS +MKNG + VLP KLR P R+L ++N Q+SGYVQGQ+ VA
 Sbjct: 185 ITVAIIISPFIIFYFLRDSSHMKNGLVNVLPKLRVPMVRVLGDINKQLSGYVQGQVTV 244
 Query: 246 ITVGVIISIMYSIIGLRYGVTLGIIAGVLNMVYPYLGSLFVAQIPVFILALVAGPVMVVKVA 305
 I VG +FSIM+S++GL+Y +T GIIAG LNM+PYLGSLF+A IPV I+A+V GP M+VKV
 Sbjct: 245 IVVGFMFSIMFSLVGLKYAITFGIAGFLNMIPYLGSLFAMIPVIMAMVQGPFMLVKVL 304
 Query: 306 IVFVIEQTLEGRFVSPLVLGNKLSIHPITIMFILLTSGAMFGVWGVFLSIPIYASIKVVV 365
 ++F+IEQT+EGRFV+PLVLGNKLSIHPITIMF+LLT+G+MFGVWGVFL IPIYAS+KVV+
 Sbjct: 305 VIFMIEQTIEGRFVAPLVGNKLSIHPITIMFLLTAGSMFGVWGVFLVIPIYASIKVVI 364
 Query: 366 KELFDWYKAVSGLYTVDV-VTEERSEEVK 393
 KELFDWYK VSGLY +V V EE + VK
 Sbjct: 365 KELFDWYKAVSGLYDEEVLVIEEVKDHVK 393

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 888

A DNA sequence (GBSx0942) was identified in *S. agalactiae* <SEQ ID 2697> which encodes the amino acid sequence <SEQ ID 2698>. Analysis of this protein sequence reveals the following:

Possible site: 58
 >>> Seems to have no N-terminal signal sequence

-978-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2715(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9891> which encodes amino acid sequence <SEQ ID 9892> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAA25160 GB:L16975 ORF1 [Lactococcus lactis]

Identities = 132/345 (38%), Positives = 203/345 (58%), Gaps = 3/345 (0%)

Query: 79 INLAQIVAEDGDIEQAFLYLDYISEDSQEYVSALLVMADLYDMEGLTDVAREKLLASKL 138
 +NLA+I ++G++++A YL I + + Y++AL+ +ADLY E + A KL A +L

Sbjct: 1 VNLAIEAEDNGNLDEALNYLIQIPVDENYIAALIKIADLYQFEVDFETAISKLEEARL 60

Query: 139 SDDPLVTFGLAEMNLSLEHYQEAIEGYASLDNREILETTGVSTYQRIGKSYAIMGKFDAA 198
 SD PL+TF LAE Y AI YA L R+IL T +S YQRIG SYA +G F+ A

Sbjct: 61 SDSPLITFALAESYFEQGDYSAAITEYAKLSERKILHETKISYQRIGDSYAQLGNFENA 120

Query: 199 IEFLEKAVDIEYDDLTFELATILYDQEEYQKANLYFKQLDTINPDFAGYEYIYGLSLRE 258
 I FLEK+++ + T++++A + + +A FK+L+ ++ +F YE Y +L

Sbjct: 121 ISFLEKSLEFDEKPEPTYKIALLYGETHNETRAIANFKRLEKMDVEFLNYELAYAQTLA 180

Query: 259 EHKSEEARLRLVQQGIRKNSFDGQQLLLASQLSYELHDVHSSSEYKQAEKVSENQDEIVM 318
 + + AL + ++G++KN LL AS++ ++L D ++E YL A + E DE V

Sbjct: 181 NQEFKAALMAKKGKMPNPNVPLLFASKICFKLKDAAAERYLVDALNLPDLHDETTF 240

Query: 319 RLSNLYLEERFEEVLELDN-DNLENILAKWNIKAHKALEMDDSDV--YYQSLYNDLKD 375
 L+NLY EE FE V+ L+ E++LAKW A AHKALE D Y + + +L +

Sbjct: 241 LLANLYFNEEDFEAVINLEELLEDEHLLAKWLPAHAHKALENDSEAAALYEELIQINLSE 300

Query: 376 NPEFLQDYAYILREFGYLDKAQEVGKAYLKLVPDDIEMSEWVNNI 420

NPEFL+DY L+E G + K + + + YL+LVPDD M + ++

Sbjct: 301 NPEFLEDYIDFLKEIGQISKTEPIEQYLELVPDDENMRNLLTDL 345

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2699> which encodes the amino acid sequence <SEQ ID 2700>. Analysis of this protein sequence reveals the following:

Possible site: 25

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2991(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 267/409 (65%), Positives = 336/409 (81%), Gaps = 1/409 (0%)

Query: 13 MLNSEKMIVSIQNDLEHANKYFEKALKNDPEEVLLLELGAYLESIGFLPQAKRLYDQIRP 72

MLNSEKMI S+ QDL HA KYF+KALK D + L+ LG YLESIGFLP AKR+Y Q+

Sbjct: 7 MLNSEKMIASLDQQLAHAEKYFQKALKEDDADSLIALGEYLESIGFLPHAKRTYLQLAD 66

Query: 73 NYPEVAINLAQIVAEDGDIEQAFLYLDYISEDSQEYVSALLVMADLYDMEGLTDVAREKL 132
 +YPE+ INLAQI AED IE+AFLYLD +S+DS Y+SALLVMADLYDMEGLT+VAREKL

Sbjct: 67 DYPELNINLAQIAAEDDAIEBAFLYLDKVS KDSPNYLSALLVMADLYDMEGLTEVAREKL 126

Query: 133 LLASKLSDDPLVTFGLAEMNLSLEHYQEAIEGYASLDNREILETTGVSTYQRIGKSYAIM 192
 L A +S +PLV FGLAE+++SL+H++EAI+ YA LDNR+ILE TG+STYQRIG++YA +

Sbjct: 127 LQAVGISPEPLVIFGLAIDMSLQHFKEAIDYYAQLDNRIQLTGTISTYQRIGRAYASL 186

Query: 193 GKFDAAIEFLEKAVDIEYDDLTFELATILYDQEEYQKANLYFKQLDTINPDFAGYEYIY 252
 GKF+AAIEFLEKAV IEY+D TVFELAT++YDQE YQKANLYFKQL+TINPD+ GY EY Y

-979-

Sbjct: 187 GKFEAAIEFLEKAVAIEYEDETVELATLMYDQENYQKANLYFKQLETINPDYPGYEYGY 246

Query: 253 GLSLREEHKSEELRLVQQGIRKNSFDGQLLLASQLSYELHDVHSSSESYLKQAEKVSEN 312
LSL EEHK+ EALRLVQQG+RKN+FD QLLLLASQLSYELHD ++E+YL QA++V+ +

5 Sbjct: 247 ALSLHEEHKTSEALRLVQQGLRKNAFDSQLLLASQLSYELHQRQNAENYLLQAKEVAVD 306

Query: 313 QDEIVMRLSNLYLEERFEEVLELDNDNLENILAKWNIKAKHAKALEMDD-SVDYYQSLYN 371
+EI+MRL LY + ERFEEV+ L+ + ++N+L KW IAKA+ ALE ++ ++ Y +

10 Sbjct: 307 DEEILMRLVTLYFDAERFEEVIALNRETIDNVLTQWTTAKAYHALEQEEVALALYNEISA 366

Query: 372 DLKDNPEFLQDYAYILREFGYLDKAQEVGKAYLKLVPDDIEMSEWVNNI 420
DL +NPEFLQDYAY+LREFG KA ++ AYL+ VPDD+ M ++++I

Sbjct: 367 DLAEENPEFLQDYAYLLREFGQFHKAIQMATAYLRQVPDDVNMQDFLDHI 415

15 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 889

A DNA sequence (GBSx0943) was identified in *S.galactiae* <SEQ ID 2701> which encodes the amino acid sequence <SEQ ID 2702>. This protein is predicted to be alpha-acetolactate synthase (ilvK). Analysis of this protein sequence reveals the following:

Possible site: 60
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

25 bacterial cytoplasm --- Certainty=0.2105(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

30 >GP:CAA01700 GB:A23961 alpha-acetolactate synthase [Lactococcus
lactis]
Identities = 396/559 (70%), Positives = 466/559 (82%), Gaps = 8/559 (1%)

Query: 4 SHNQYGADLIVDSLINHDKVYVFGIPGAKIDRVFDTLE-DKGPFLIVARHEQNATFMAQA 62
S Q+GA+L+VDSLINH VKYVFGIPGAKIDRVFD LE ++GF+++V RHEQ A FMAQA

35 Sbjct: 2 SEKQFGANLVVDSLINHVKYVFGIPGAKIDRVFDLLENEEGPQMVVTRHEOGAAAFMAQA 61

Query: 63 VGRITGEPGVVIATSGPGISNLATGLVTATDEGDAVLAIGGQVKRGDILLKRAHQSMNVA 122
VGR+TGEPEGVV+ TSGPG+SNLAT L+TAT EGDA+LAIGGQVKR D LKRAHQSM+N

40 Sbjct: 62 VGRITGEPGVVVITSGPGVSNLATPLLTATSEGDAILAIGGQVKRSDRLKRAHQSMNDAG 121

Query: 123 MLEPITKYSAEVHDPNTLSETVANAYRLAKSGKPGASFISIPQDVTDSPVSVKAIKPLSA 182
M++ TKYSAEV DPNTLSE++ANAYR+AKSG PGA+F+SIPQDVTD+ VS+KAI+PLS

45 Sbjct: 122 MMQSATKYSAEVLDPNTLSESIANAYRIAKSGHGPATFLSIPQDVTDAEVSIAIKAIPLSD 181

Query: 183 PKLGSASVLDINYLQAQAINNAVLPLVLLGNGASSEGVTAAVRRLLDVAVKLPVVETFCQAG 242
PK+G+AS+ DINYLAQAI NAVLPV+L+G GAS V +++R LL V +PVVETFCQAG

Sbjct: 182 PKMGNASIDDINYLQAQIKNAVLPLVILVGAGASDAKVASSLRNLLTHVNIPVVETFCQAG 241

50 Query: 243 IVSRELEDETFFGVRVGLFRNQPGDMLLKRADLVIAIGYDPIEYARNWNAEISARIIVID 302
++S +LE TF+GR+GLFRNQPGDMLLKR+DLVIA+GYDPIEYARNWNAEI +RIIVID

Sbjct: 242 VISHDLE-HTFYGRIGLFRNQPGDMLLKRSIDLVIAGYDPIEYARNWNAEIDSRIIVID 300

Query: 303 VEQAIEDTYFQPERELIGDMAHTLDLLPAIKGYELPEGSKEYLKLGRNNIENVSVDVKFD 362
AEIDTY+QPERELIGD+A TLD LLP++GY++P+G+K+YL GL E +FD

55 Sbjct: 301 NALAEIDTYFQPERELIGDIAATLDNLLPAVRGYKIPKGT KDYLDGLH---EVAEQHEFD 357

Query: 363 RDSA-HGLVHPLDLIDVLQENTTDDMTVTVDVGSYIWMARYFKSYEARHLLFSNGMQTL 421
++ G +HPLDL+ QE DD TVTVDVGS YIWMAR+FKSYE RHLLFSNGMQTL

60 Sbjct: 358 TENTEEGRMHPLDLVSTFQEIIVKDETVTVDVGSYIWMARHFKSYEPRHLLFSNGMQTL 417

-980-

Query: 422 GVALPWAI SAALLRPNTKVISVSGDGGFLFSAQELETAVRLHLPIVHIIWNDGKYNMVEF 481
 GVALPWAI+AALLRP KV S SGDGGFLF+ QELETAVRL+LPIV I IWNDG Y+MV+F
 Sbjct: 418 GVALPWAITAALLRPGKVVYSHSGDGGFLFTGQELETAVRLNLPVQIIWNDGHYDMVKF 477

5 Query: 482 QEEMKYGRSSGVDFGPVDFVVKYAESFGAKGYRVDSKDSFEETLKQALIDAENG PVLIDVP 541
 QEEMKYGRS+ VDFG VD+VKYAE+ AKGYR SK+ E LK I GPV+IDVP
 Sbjct: 478 QEEMKYGRSAVDFGYVDYVVKYAEAMRAKGYRAHSKEELAEILKS--IPDTTGPVVIDVP 535

10 Query: 542 IDYKDNVTLGETILPDEFY 560
 +DY DN+ L E +LP+EFY
 Sbjct: 536 LDYSDNIKLAEKLLPEEFY 554

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 890

A DNA sequence (GBSx0944) was identified in *S.galactiae* <SEQ ID 2703> which encodes the amino acid sequence <SEQ ID 2704>. This protein is predicted to be alpha-acetolactate decarboxylase (aldC). Analysis of this protein sequence reveals the following:

20 Possible site: 43
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 25 bacterial cytoplasm --- Certainty=0.3096(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9889> which encodes amino acid sequence <SEQ ID 9890> was also identified.

30 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA57941 GB:X82620 alpha-acetolactate decarboxylase [Lactococcus lactis]
 Identities = 139/239 (58%), Positives = 187/239 (78%), Gaps = 3/239 (1%)

35 Query: 16 MSETVKLPQYSTLSSLMAGLYKGS LTIGELLTHGDLGIGTVH MIDGELIVLDGKAYQAIG 75
 MSE +LPQY+TL +LMAGLY+G++TIGELL HGDLGIGT+ IDGELIVLDGKAYQA
 Sbjct: 1 MSEITQLFQYNTLGALMAGLYEGTMTIGELLKHGDLGIGTLD SIDGELIVLDGKAYQA-- 58

40 Query: 76 TDGKAEIIQLSDDVTVPYAAVLPHHIQKQFDINAEIDNKDLEEMILKNFEGQNLFKSLKI 135
 G I++L+DD+ VPYAAV+PH + F + +K+LE+ I F+GQNLFS+KI
 Sbjct: 59 -KGDKTIVELTDDIKVPYAAVPHQAEVVFQKQFTVSDKELEDRIESYFDGQNLFRS IKI 117

45 Query: 136 KGTFSRMHVRMIPKSPQHKRFADIASNQPEFTRENVSGTLVGIWTPELFHG VGVKGFEHVH 195
 G F +MHVRMIP++ +F +++ NQPE+T EN+ GT+VGIWTP E+FHGV V G+H+H
 Sbjct: 118 TGKFKPMHVRMIPRAKSGTKFVEVSNQPEYTEENIKGTIVGIWTP EMFHGVSVAGYHLH 177

50 Query: 196 FISDDLTFGGHVM DYSLTQGVKVEIGKVDQLDQCFTQDQEF LKANFDLQKLREDIDLSE 254
 FIS+D TFGGHV+D+ + G VEIG +DQL+Q FP QD++FL A+ D++ L++DID++E
 Sbjct: 178 FISEDFTFGGHVLDFTIINDGTVEIG AIDQLNQSFVQDRKFLFADLDIEALKKDIDVAE 236

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 891

A DNA sequence (GBSx0945) was identified in *S.agalactiae* <SEQ ID 2705> which encodes the amino acid sequence <SEQ ID 2706>. This protein is predicted to be fibronectin-binding protein-like protein A. Analysis of this protein sequence reveals the following:

5 Possible site: 57
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----

10 bacterial cytoplasm --- Certainty=0.5042(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

15 >GP:CAA46282 GB:X65164 fibronectin-binding protein-like protein A
 [Streptococcus gordonii]
 Identities = 392/550 (71%), Positives = 462/550 (83%)

 Query: 1 MSFDGFFLHHLTNELQEIEKGRIQKVNQPFDELVLTRNNRRNYKLLLSAHPVFGRIQ 60
 MSFDGFFLHH+T EL+ ++ GRIQK+NQPF+ ELVL IR+NR++ KLLLSAH VFGR+Q
 20 Sbjct: 1 MSFDGFFLHHMTEELRHLELVGGRIQKINQPFQELVLQIRSNRKSLLLSAHSVFGRVQ 60

 Query: 61 TTEANFQNPQNPNTFTMIMRKYLGAVIETIQIENDRILEIVVSNKNEIGDHIKATLVV 120
 T+ F+NP PNTF M-MRKYLQGAVIE IQQ+ENDRILEI VSNKNEIGD + TLV+
 25 Sbjct: 61 LTDTTFENPAVPNTFIMVMRKYLGAVIEAIQQVENDRILEISVSNKNEIGDSVAVTLVI 120

 Query: 121 EIMGKHSNIILIDKNEHKIIESIKHVGFSQNSYRTILPGSTYIAPPKTKAINPFDISDQT 180
 EIMGKHSNIIL+DK KIIE+IKHVGFSQNSYRTILPGSTY+APP+T ++NPF + D+
 30 Sbjct: 121 EIMGKHSNIILLDKASGKIIEAIKHVGFSQNSYRTILPGSTYVAPPQTGSLNPFTVGDEK 180

 Query: 181 LFEILLQTNLSPKNLQQLLQGLGRDTALELSHCLKDNKLNDFRQFFSREYYPSLTEKSFS 240
 LFE+LQT ++ PK L Q+ QGLGRDTA ELS L ++L FR FF+ PSLTEKSFS
 35 Sbjct: 181 LFEILLQTEIEPKRLQLIFQGLGRDTATELSGRLTTDRLKTPRAFFASPTQPSLTEKSFS 240

 Query: 241 AVQFSSSHETFQSLGQLLDYYYQEKAEEKDRIAQQASDLIHRVQSELEKNIKKLAKQQDEL 300
 A+ FS S +L +LLD +Y++KAE+ R+ QQAS+LI RV++ELEKN KKL KQ+DEL
 40 Sbjct: 241 ALVFSDSKTQMSTLSELDDTFYKDKAERYRVNQASELIRRVENELEKNRKKLGKQCEDEL 300

 Query: 301 LATENAEERFQKGELLTYLSMVPNNQDVVLDNYTNTQIEISLDRALTPNQNAQRYFK 360
 LATE AEEFRQKGELLTT+L VPN+QD V LDNYT + I I+LD+ALTPNQNAQRYFK
 45 Sbjct: 301 LATEKAEERFQKGELLTTFLHQPNDQDQVELDNYTGEKILITLDKALTPNQNAQRYFK 360

 Query: 361 KYQKLKEAVKHLKGIISDENTITTYLESVETSLNHASMEDINDIREELVETGFIKRRRAHD 420
 +YQKLKEAVKHL +I +T TI YLESVET+L AS+ +I +IREEL++TGFI+RR +
 50 Sbjct: 361 RYQKLKEAVKHLTSLIEETRITTYLESVETALAQAASLTEIAEIREBELIQTGFIRRRQRE 420

 Query: 421 KQHKRKKPEQYLASDGKTIIMVGRNNLQNDLTFKMARKGELWFAKDIPGSHVLIRDNL 480
 K KRRKPE+YLASDG+TII+VGRNNLQNDLTFKMA+K ELWFAKDIPGSHV+I NL
 55 Sbjct: 421 KIQRKKRKKPEKYLASDGQTIILVGRNNLQNDLTFKMAKDELWFAKDIPGSHVVITGNL 480

 Query: 481 NPSDEVKTDAAELAAYYSKARLSNLVQVDMIEAKKLNKPSGTPGPFVTTYTGQKTLRVTP 540
 PSDEVKTDAAELAAY+SKARLSNLVQVDMIE KKLNKP+G KPGFVTTYTGQKTLRVTP
 60 Sbjct: 481 QPSDEVKTDAAELAAYFSKARLSNLVQVDMIEIKKLNKPTGGKPGFVTTYTGQKTLRVTPD 540

 Query: 541 QEKIDSLKLK 550
 +KI S+K++
 Sbjct: 541 ADKIKSMKIQ 550

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2707> which encodes the amino acid sequence <SEQ ID 2708>. Analysis of this protein sequence reveals the following:

60 Possible site: 38

 >>> Seems to have no N-terminal signal sequence

-982-

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.5434(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein differs significantly from L28919 in its mid-region:

10 Query: 223 QHFQGLGRDTAKELAEELLTTD
 F L +T K + ELLTTD
 Sbjct: 121 PAFSRLRGETPKRIGELLTTD

An alignment of the GAS and GBS proteins is shown below.

15 Identities = 421/549 (76%), Positives = 487/549 (88%)

Query: 1 MSFDGFFLHHLTNELQEIEKGRIQKVNQPPFDHELVLTI RNRRNYKLLLSAHPVFGRIO 60
 MSFDGFFLHHLTNEL+E + GRIQKVNQPF+ ELVLTI RN+R+NYKLLLSAHPVFGR+Q
 Sbjct: 27 MSFDGFFLHHLTNELKENLLYGRIQKVNQPFERELVLTI RNHRKNYKLLLSAHPVFGRVQ 86

20 Query: 61 TTEANFQNFQNPNTFTMIMRKYLGAVIETIQIENDRIEIVVSNKNEIGDHIKATLVV 120
 T+A+FNQFQ PNTFTMIMRKYLGAVIE ++QI+NDRI+EI VSNKNEIGD I+ATL++
 Sbjct: 87 ITQADFQNFQVPNTFTMIMRKYLGAVIEQLEQIDNDRIIEIKVSNKNEIGDAIQATLII 146

25 Query: 121 EIMGKHSNIILIDKNEHKIIESIKHVGFQNSYRTILPGSTYIAPPKTKAINPFDISDQT 180
 EIMGKHSNIIL+D+ E+KIIESIKHVGFQNSYRTILPGSTYI PPKT A+NPF I+D
 Sbjct: 147 EIMGKHSNIILVDRAENKIIESIKHVGFQNSYRTILPGSTYIEPPKTA AVNPFTITDVP 206

30 Query: 181 LFEELLQTNDLSPKNLQQLLQGLGRDTALELSHCLKDNKLNDFRQFFSREYYP SLTEKSFS 240
 LFE+LQT +L+ K+LQQ QGLGRDTA EL+ L +KL FR+FF+R +LT SF+
 Sbjct: 207 LFEILQTQELTVKSLQHFQGLGRDTAKELAEELLTTDKLKRFEFFARPTQANLTTASFA 266

35 Query: 241 AVQFSSSHETFQSLGQLLDYYYQEKA EKDRIAQQASDLIHRVQSELEKNIKKLAKQQDEL 300
 V FS SH TF++L +LD++YQ+KAE+DRI QQASDLIHRVQ+EL+KN KL+KQ+ EL
 Sbjct: 267 PVLFSDSHATFETLSMDLDFYQDKAERDRINQQASDLIHRVQTELDKNRNKLKQEAEL 326

40 Query: 301 LATENAEFRQKGELLTTYLSMVPNNQDVVLDNYTNTQTI EISLDRALTPNQNAQRYFK 360
 LATENAE FRQKGELLTTYLS+VPNNQD V+LDNYT + IEI+LD+ALTPNQNAQRYFK
 Sbjct: 327 LATENAEFRQKGELLTTYLSLVPNNQDSVILDNYTGEKTEIALKALTPNQNAQRYFK 386

45 Query: 361 KYQKLKEAVKHLKGIISDTE NTITYLESVETSLNHASMEDINDIREELVETGFIKRAHD 420
 KYQKLKEAVKHL G+I+DT+ +ITY ES+ +L+ AS++DI DIREEL + GF+K R D
 Sbjct: 387 KYQKLKEAVKHLGLIADTKQSITYFESVDYNLSQASIDDIEDIREELYQAGFLKSRQD 446

50 Query: 421 KQHKRKKPEQYLASDGKTIIMVGRNNLQNDLTFKMARKGELWFWHAKDIPGSHVLI RDNL 480
 K+HKRKKPEQYLASDG TI+MVGRNNLQN+ELTFKMA+KGELWFWHAKDIPGSHV+I+DNL
 Sbjct: 447 KRHKRKKPEQYLASDGTIIMVGRNNLQNEELTFKMAKKGELWFWHAKDIPGSHVII RDNL 506

55 Query: 481 NPSDEVKTDAAELAAYYSKARLSNLVQVDMIEAKKLKPSGKPGFVTTYTGQKTLRVTP 540
 +PSDEVKTDAAELAAYYSKARLSNLVQVDMIEAKKL+KPSG KPGFVTTYTGQKTLRVTP
 Sbjct: 507 DPSDEVKTDAAELAAYYSKARLSNLVQVDMIEAKKLKPSGAKPGFVTTYTGQKTLRVTPD 566

 Query: 541 QEKIDSLKL 549
 Q KI S+KL
 Sbjct: 567 QAKILSMKL 575

SEQ ID 2706 (GBS81) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 6 (lane 2; MW 64kDa) and in Figure 6 (lane 5; MW 64kDa). The GBS81-His fusion product was purified (Figure 190, lane 3) and used to immunise mice. The resulting antiserum was used for FACS (Figure 319), which confirmed that the protein is immunoaccessible on GBS bacteria.

60 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-983-

Example 892

A DNA sequence (GBSx0946) was identified in *S.galactiae* <SEQ ID 2709> which encodes the amino acid sequence <SEQ ID 2710>. Analysis of this protein sequence reveals the following:

```

Possible site: 53
5  >>> Seems to have an uncleavable N-term signal seq
    INTEGRAL    Likelihood = -9.08    Transmembrane    6 - 22 ( 1 - 24)

----- Final Results -----
10      bacterial membrane --- Certainty=0.4630(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15  >GP:AAF94260 GB:AE004191 conserved hypothetical protein [Vibrio cholerae]
    Identities = 111/295 (37%), Positives = 184/295 (61%), Gaps = 1/295 (0%)

Query: 36  QVVKIGILQYVTHDALDAIEKGVEDGLAQEGYK-GKKVKLTVLNAEADQSKIQAMSKQLV 94
      + K+ + Q V H ALDA +G+ DGL +GY+ GK ++ A+ + + +++Q V
20  Sbjct: 26  KTAKVAVSQIVEHPALDATRQGLLDGLKAKGYEKGKLEFDYKTAQGNPAIAVQIARQFV 85

Query: 95  NHHNDILIGIATPSAQGLAASTKDTPIIMGAVSDPLGAKLVTNMKKPTTNVTGLSNVVP 154
      + D+L+GIATP+AQ L ++TK PI+ AV+DP+GAKLV +++P NVTGLS++ P
25  Sbjct: 86  GENPDVLVGIATPTAQALVSATKTIPIVFTAVTDPVGAKLVKQLEQPGKNVTGLSDLSPV 145

Query: 155 KQTVOLIKDITPNIKRIGILYASSEDNSVSQVTEFTKYAQKAGLEVLYKYSVPSTNEIKTS 214
      +Q V+LIK+I PN+K IG++Y E N+VS + A K G+++++ + +++++
30  Sbjct: 146 EQHVELIKEILPNVKSIGVVYNPGEANAVSLMELLKLSAAKHGIKLVEATALKSADVQSA 205

Query: 215 MSVMTKKVDVAVFVPQDNTIASAFRTVIVAANQANIPVYSSVDTMVEQGSIASVAQSQYGL 274
      + +K D ++ DNT+ASA +IVAANQA PV+ + + VE+G+IAS+ Y +
35  Sbjct: 206 TQATAEKSQVIYALIDNTIVASAIKGMIVAANQAKTPVFGAATSYVERGATIASLGFDDYYQI 265

Query: 275 GLETAKQAIKVLRGKPKVDVPVKVIDTGKPSLNLKAAKHLGIKIPKKIMKQAEIT 329
      G++TA +L GK + V+V +N AA+ LGI IP+ ++ +A T
40  Sbjct: 266 GVQTADYVAAILEGKEPGSLDVQVAKGSDLVINKTAAEQLGITIPAVLARATST 320

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2711> which encodes the amino acid sequence <SEQ ID 2712>. Analysis of this protein sequence reveals the following:

```

Possible site: 23
40  >>> Seems to have an uncleavable N-term signal seq
    INTEGRAL    Likelihood = -11.25    Transmembrane    6 - 22 ( 1 - 27)

----- Final Results -----
45      bacterial membrane --- Certainty=0.5501(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

50  >GP:AAF94260 GB:AE004191 conserved hypothetical protein [Vibrio cholerae]
    Identities = 103/304 (33%), Positives = 178/304 (57%), Gaps = 1/304 (0%)

Query: 17  VIGSLLSKGVSKENRDLANQQNITIGILQFVTHEALDDIKRGIEDQLK-KQMPQKQNVVI 75
      VI + + G + + + + + Q V H ALD ++G+ D LK K + +N+
55  Sbjct: 6  VIATAVLAGAALLSSQSISMAKTAKVAVSQIVEHPALDATRQGLLDGLKAKGYEKGKLEF 65

Query: 76  KVMNAEGDQSKIQTMSRQLVQSGSDIVIGIATPAAQGLAATSKDIPVMSAVSDPVGSRL 135
      A+G+ + ++RQ V D+++GIATP AQ L + +K IP+V +AV+DPVG++L
60  Sbjct: 66 DYKTAQGNPAIAVQIARQFVGENPDVLVGIATPTAQALVSATKTIPIVFTAVTDPVGAKL 125

Query: 136 VMQLDQPEANVTGLSNKVPVKQTIDLMKKLTPHVKTVGILYASNEDNSLSQVKEFRRLAR 195

```

-984-

V QL+QP NVTGLS+ PV+Q ++L+K++ P+VK++G++Y E N++S ++ + A
 Sbjct: 126 VKQLEQPGKNVTGLSDLSPVEQHVELIKEILPNVKSIGVVYNPGEANAVSLMELLLKLSAA 185
 Query: 196 KKGYSYVAVPSTNEVPATMSVMLGKVDVFIPODNTIASAFSSVMTTSSKAAKIPVYTS 255
 K G +++ + +V + + K D ++ DNT+ASA ++ + AK PV+ +
 Sbjct: 186 KHGIXLVEATALKSADVQSATQAIAEKSDVIYALIDNTIVASAIEGMIVAAANQAKTPVFGA 245
 Query: 256 VDRMVEKGGGLAISQNYDLGVQTANQVLKLIKGRVVDVPVKVVDIGQPLINKNVAAEL 315
 VE+G +A++ + Y +GVQTA+ V +++GK + V+V +INK A +L
 Sbjct: 246 ATSYVERGAIASLGFDYQIGVQTADYVAAILEGKEPGSLDVQVAKGSDLVINKTAAEQ 305
 Query: 316 GIAI 319
 GI I
 Sbjct: 306 GITI 309

An alignment of the GAS and GBS proteins is shown below.

Identities = 181/322 (56%), Positives = 252/322 (78%), Gaps = 1/322 (0%)
 Query: 1 MKNKGLIATLILTLVVGELFYNK-SEKRLNLSEKQVVKIGILQYVTHDALDAIEKGVE 59
 MKNK LIATL++LT++V+G L S++ +L+ +Q + IGILQ+VTH+ALD I++G+E
 Sbjct: 1 MKNKSLIATLLVLTIVIVIGSLLSKGVSKENRDLANQQNITIGILQFVTHEALDDIKRGIE 60
 Query: 60 DGLAQEGYKGGKVKLTFLVNAEADQSKIQAMSKQLVNHNDILIGIATPSAQGLAASTKDT 119
 D L ++ + + V + V+NAE DQSKIQ MS+QLV +DI+IGIATP+AQGLAA++KD
 Sbjct: 61 DQLKKMPQKQNVVIVKMNAEGDQSKIQTMSRQLVQSGSDIVIGIATPAAQGLAATSKDI 120
 Query: 120 PIIMGAVSDPLGAKLVNMMKPTTNTVGLSNVPTKQTVOLIKDITPNIKRIGILYASSE 179
 P++M AVSDP+G++LV + +P NVTGLSN VP KQT+ L+K +TP++K +GILYAS+E
 Sbjct: 121 PVVMSAVSDPVGSRRLVMQLDQPEANVTGLSNKVPVKQTIDLMKKLTPHVKTGVGILYASNE 180
 Query: 180 DNSVSQVTEFTKYAQKAGLEVLYKYSVPSTNEIKTSMVMTKKVDAVFVPQDNTIASAFRT 239
 DNS+SQV EF + A+K G +V+ Y+VPSTNE+ +MSVM KVDVAF+PQDNTIASAF +
 Sbjct: 181 DNSLSQVKEFRRLARKKGYQVYSYAVPSTNEVPATMSVMLGKVDVFIPODNTIASAFSS 240
 Query: 240 VIVAANQANIPVYSSVDTMVEQGSIASVAQSQYGLGLETAQQAIVLRGKPVKDVVPVKVI 299
 V+ + A IPVY+SVD MVE+G +A+++Q+QY LG++TA Q +K+++GK V DVPVKV+
 Sbjct: 241 VMTTSSKAAKIPVYTSVDRMVEKGGGLAISQNYDLGVQTANQVLKLIKGRVVDVPVKV 300
 Query: 300 DTGKPSLNLKAAKHLGKIPKK 321
 D G+P +N A LGI I K+
 Sbjct: 301 DIGQPLINKNVAAELGIAIKKE 322

SEQ ID 2710 (GBS254) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 43 (lane 4; MW 27kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 48 (lane 3; MW 59.6kDa).

GBS254-GST was purified as shown in Figure 203, lane 6.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 893

A DNA sequence (GBSx0947) was identified in *S.agalactiae* <SEQ ID 2713> which encodes the amino acid sequence <SEQ ID 2714>. This protein is predicted to be probable permease of ABC transporter (rbsC). Analysis of this protein sequence reveals the following:

Possible site: 24
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -15.12 Transmembrane 127 - 143 (119 - 151)
 INTEGRAL Likelihood = -8.81 Transmembrane 206 - 222 (200 - 227)
 INTEGRAL Likelihood = -6.48 Transmembrane 260 - 276 (258 - 282)
 INTEGRAL Likelihood = -5.84 Transmembrane 234 - 250 (231 - 257)

-985-

```

INTEGRAL    Likelihood = -4.78    Transmembrane    55 - 71 ( 54 - 72)
INTEGRAL    Likelihood = -3.61    Transmembrane    177 - 193 ( 176 - 194)
INTEGRAL    Likelihood = -3.35    Transmembrane    84 - 100 ( 83 - 102)
INTEGRAL    Likelihood = -1.91    Transmembrane    10 - 26 ( 10 - 26)

```

----- Final Results -----

```

bacterial membrane --- Certainty=0.7050(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAG07224 GB:AE004801 probable permease of ABC transporter
[Pseudomonas aeruginosa]

Identities = 116/288 (40%), Positives = 185/288 (63%), Gaps = 9/288 (3%)

```

Query: 2    IISSVSQGLLWGILGLGIYLTFRILKFPDMTTEGSFPLGGAVCVTLMNQGVNPILATILG 61
+ ++ GL++ ++ LG++++FR+L+FPD+T +GSFPLGGAVC TL+ G +P AT+
Sbjct: 6    LFGALEIGLIFSLVALGVFISFRLLRFPDLTVDGSPFLGGAVCATLIALGWDYPYSATLAA 65

```

```

Query: 62   MLSGMLAGFVTGLLYTKGKIPTILAGILVMTSCHSIMLMVMKRANLGLNEIQTLDKDFLPF 121
+G LAG TGLL K KI +LA IL+M + +SI L +M + N+ L TL L
Sbjct: 66   TAAGALAGLATGLLNVLKIMDLLASILMMIALYSINLRIMGKPNVPLIAEPTLFTLLQP 125

```

```

Query: 122  SNDLNLVLGLIAILLVISA---LIYFLYTRLGQAYIATGDNPDMAKSGIDTDKMEMILG 178
+ + L+ + +VI+A L +F T+ G A ATG NP MA++ G++T M +LG
Sbjct: 126  EWLSDYVFRPLLVLVFIIVIAAKLLLDWFFTTQKGLAIRATGSNPRMARAGQVNTGGMILLG 185

```

```

Query: 179  LIVSNGLIALSGALVVSQQDGYADVSKGIGVIVIGLASIIIGE-VLYSTGLTLFERLIAIV 237
+ +SN L+AL+GAL +Q G AD+S GIG IVIGLA++I+GE +L S L L +A++
Sbjct: 186  MAISNALVALAGALFAQTGGADISMGIGTIVIGLAAVIVGESILPSRRLIL--ATLAVI 243

```

```

Query: 238  VGSILYQFLITAVI---ALGFNTNYLKLFSIAIVLIGICLMVPVLKTKIL 282
+G+I+Y+F I + +G L L +A+++ + L++P++K ++L
Sbjct: 244  LGAIVYRFFIALALNSDFIGLQAQDLNLVTAVLVTVALVIPMMKKRL 291

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2715> which encodes the amino acid sequence <SEQ ID 2716>. Analysis of this protein sequence reveals the following:

Possible site: 55

>>> Seems to have an uncleavable N-term signal seq

```

INTEGRAL    Likelihood = -10.46    Transmembrane    131 - 147 ( 125 - 156)
INTEGRAL    Likelihood = -8.65     Transmembrane    210 - 226 ( 204 - 230)
INTEGRAL    Likelihood = -8.17     Transmembrane    265 - 281 ( 261 - 283)
INTEGRAL    Likelihood = -7.22     Transmembrane    238 - 254 ( 233 - 261)
INTEGRAL    Likelihood = -3.03     Transmembrane    89 - 105 ( 87 - 107)
INTEGRAL    Likelihood = -2.60     Transmembrane    63 - 79 ( 62 - 79)
INTEGRAL    Likelihood = -2.23     Transmembrane    180 - 196 ( 180 - 198)
INTEGRAL    Likelihood = -2.13     Transmembrane    14 - 30 ( 14 - 30)

```

----- Final Results -----

```

bacterial membrane --- Certainty=0.5182(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

>GP:AAG07224 GB:AE004801 probable permease of ABC transporter
[Pseudomonas aeruginosa]

Identities = 118/285 (41%), Positives = 186/285 (64%), Gaps = 7/285 (2%)

```

Query: 6    IISSVSQGLIHWGLGLGIYLTFRILNFPDMTTEGSFPLGGAVAVTAISLGWNPFLSTLLG 65
+ ++ GLI+ ++ LG++++FR+L FPD+T +GSFPLGGAV T I+LGW+P+ +TL
Sbjct: 6    LFGALEIGLIFSLVALGVFISFRLLRFPDLTVDGSPFLGGAVCATLIALGWDYPYSATLAA 65

```

```

Query: 66   MLSGALAGFLTGLLYTKGKMPDLLAGILVMTSCNSIMLMVMGRANLGLHDHKRIQDCPLP 125
+GALAG TGLL K K+ LLA IL+M + SI L +MG+ N+ L + L

```

-986-

Sbjct: 66 TAAGALAGLATGLLNVRKIKIMDLLASILMMIALYSINLRIMGKPNVPLIAEPTLFTLLQP 125

Query: 126 SIDLNSLLTGLITVVIVIS---VLIYFLYTNLGQAYIATGDNKDMAKSFINTDWMVMG 182
 + + L+ V IVI+ +L +F T G A ATG N MA++ G+NT M ++G

5 Sbjct: 126 EWLSDYVFRPLLLVFIVIAAKLLLDWFFTTQKGLAIRATGSNPRMARAQGVNTGGMILLG 185

Query: 183 LVVSNLSIALSGALVSQQDGYADVSKGIGVIVIGLASIIVGEVLYSTGLTLERLIAIVI 242
 + +SN+L+AL+GAL +Q G AD+S GIG IVIGLA++IVGE + + +L L A+++

10 Sbjct: 186 MAISNALVALAGALFAQTQGGADISMGITIVIGLAIVIGESILPSRRLILATL-AVIL 244

Query: 243 GSILYQFLISVVIT---LGFNTSYLKLISALVLALCLMIPVVKER 284
 G+I+Y+F I++ + +G L L++A+++ + L+IP++K+R

Sbjct: 245 GAIVYRFFIALALNSDFIGLQAQDLNLVLTAVLVTVALVIPMMKKR 289

15 An alignment of the GAS and GBS proteins is shown below.

Identities = 227/287 (79%), Positives = 259/287 (90%)

Query: 1 MIISVSQGLLWGLGLGIYLTFRILKFPDMTTEGSFPLGGAVCVTLMNQGVNPILATIL 60
 MIISVSQGL+WG+LGLGIYLTFRIL FPDMTTEGSFPLGGAV VT ++ G NP L+T+L

20 Sbjct: 5 MIISVSQGLIWLGLGIYLTFRILNFPDMTTEGSFPLGGAVAVTAISLGWNPFLLSTLL 64

Query: 61 GMLSGMLAGFVTGLLYTKGIPTILAGILVMTSCHSIMLMVMKRANLGLNEIQTLDKDFLP 120
 GMLSG LAGF+TGLLYTKGK+PT+LAGILVMTSC+SIMLMVM RANLGL++ + ++D LP

25 Sbjct: 65 GMLSGALAGFLTGLLYTKGKMPDLLAGILVMTSCNSIMLMVMKRANLGLHDHKRIQDCLP 124

Query: 121 FSNDLNLVLGLIAILLVISALYFLYTRLGQAYIATGDNPDMAKSFIDTDKMEMLGLI 180
 FS DLN L+ GLI +++VIS LIYFLYT LGQAYIATGDN DMAKSGI+TD ME++GL+

30 Sbjct: 125 FSIDLNSLLTGLITVVIVISVLIYFLYTNLGQAYIATGDNKDMAKSFINTDWMVMGLV 184

Query: 181 VSNGLIALSGALVSQQDGYADVSKGIGVIVIGLASIIIGEVLYSTGLTLFERLIAIVVGS 240
 VSN LIALSGALVSQQDGYADVSKGIGVIVIGLASII+GEVLYSTGLTL ERLIAIV+GS

35 Sbjct: 185 VSNLSIALSGALVSQQDGYADVSKGIGVIVIGLASIIVGEVLYSTGLTLERLIAIVIGS 244

Query: 241 ILYQFLITAVIALGFNTNYLKLFSAILVGLICLMVPVLTKILKGVRL 287
 ILYQFLI+ VI LGFNT+YLKL SA+VL +CLM+PV+K + KGVRL

Sbjct: 245 ILYQFLISVVITLGFNTSYLKLISALVLALCLMIPVVKERFFKGVRL 291

A related GBS gene <SEQ ID 8681> and protein <SEQ ID 8682> were also identified. Analysis of this protein sequence reveals the following:

40 Lipop: Possible site: -1 Crend: 0
 McG: Discrim Score: 4.24
 GvH: Signal Score (-7.5): -6.43
 Possible site: 24

>>> Seems to have an uncleavable N-term signal seq

45 ALOM program count: 8 value: -15.12 threshold: 0.0

INTEGRAL	Likelihood = -15.12	Transmembrane	127 - 143 (119 - 151)
INTEGRAL	Likelihood = -7.54	Transmembrane	206 - 222 (201 - 225)
INTEGRAL	Likelihood = -6.48	Transmembrane	260 - 276 (258 - 282)
INTEGRAL	Likelihood = -5.84	Transmembrane	234 - 250 (231 - 257)
INTEGRAL	Likelihood = -4.78	Transmembrane	55 - 71 (54 - 72)
INTEGRAL	Likelihood = -3.61	Transmembrane	177 - 193 (176 - 194)
INTEGRAL	Likelihood = -3.35	Transmembrane	84 - 100 (83 - 102)
INTEGRAL	Likelihood = -1.91	Transmembrane	10 - 26 (10 - 26)
PERIPHERAL	Likelihood = 4.77		36

55 modified ALOM score: 3.52

*** Reasoning Step: 3

----- Final Results -----

60 bacterial membrane --- Certainty=0.7050(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

-987-

[illegible]

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

40 **Example 894**

A DNA sequence (GBSx0948) was identified in *S.agalactiae* <SEQ ID 2717> which encodes the amino acid sequence <SEQ ID 2718>. This protein is predicted to be ABC transporter (potA). Analysis of this protein sequence reveals the following:

```

45      Possible site: 36
      >>> Seems to have an uncleavable N-term signal seq

      ----- Final Results -----
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
50      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 9887> which encodes amino acid sequence <SEQ ID 9888> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

55 >GP:AAF86640 GB:AF162694 ABC transporter [Enterococcus gallinarum]
Identities = 117/252 (46%), Positives = 167/252 (65%)

Query: 19 MVMKIIEELKEATVQVSNGLAEMKTI LDHVNLSIYEHDFTITLGGNGAGKSTLFNVIAGTL 78
M ++ + + G +L ++L++ DFITI+GNGAGKSTL N IAGT+

-988-

Sbjct: 1 MTPVLTISDLHQTFEKGITINENHVLRGIDLTMSGDFITIIGNGAGKSTLINSIAGTI 60

Query: 79 MLSSGNIYIMGQDVTNLSAEKRAKYLRSRVFQDPKMGTA PRMTVAENLLVAKFRGEKRPLV 138
G I + +++T S +R+K +SRVFQDP+MGTA R+TV ENL +A RG+ R

5 Sbjct: 61 PTEQKIVLGDKEITRHSVTRRSKEISRVFQDPRMGTA VRILTVEENLALAYKRGQVRGFS 120

Query: 139 PRKIINYTEEFQKLIARTGNGLDRHLETPTGLSSGGQRQALSLLMATLKKPNLLLLDEHT 198
+ F++ +AR GL+ L T GLSSGGQRQA++LLMATL++P L+LLDEHT

10 Sbjct: 121 SGVKGKHAFFKEKLARLNLGLENLRLTTEIGLSSGGQRQAITLLMATLQQPKLILLDEHT 180

Query: 199 AALDPRTSVSLMGLTDEFIKQDSL TALMITHH MEDALKYGNRVLMKDGKIVRDLNQAQK 258
AALDP+TS+++M LTD+ I++ LTA M+TH MEDA++YGNR++++ GKIV D+ +K

Sbjct: 181 AALDPKTSMTVMALTDQLIQEQQLTAFMVTHDMEDAIRYGNRLIMLHQKIVVDITGEEK 240

15 Query: 259 NKMAIADYYQLF 270
+ + D LF

Sbjct: 241 QSLTVPDLMALF 252

20 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2719> which encodes the amino acid sequence <SEQ ID 2720>. Analysis of this protein sequence reveals the following:

Possible site: 58
>>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.2249(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

30 Identities = 186/250 (74%), Positives = 210/250 (83%)

Query: 22 KIIELKEATVQVSNGLAEMKTILDHVNLSIYEHDFITILGGNGAGKSTLFNVIAGTLMLS 81
KIIEL ATV V NG + KTILD+V L+IYEHDF+TILGGNGAGKSTLFNVIAGTL L+

35 Sbjct: 3 KIIELINATVDVDNGFEDAKTILDNVTLTIYEHDFITILGGNGAGKSTLFNVIAGTLSLT 62

Query: 82 SGNIIYIMGQDVTNLSAEKRAKYLRSRVFQDPKMGTA PRMTVAENLLVAKFRGEKRPLVPRK 141
G I I+GQDVT+ AEKRA YLSRVFQD KMGTA PRMTVAENLL+A+ RG KR L RK

Sbjct: 63 RGQIRILGQDVTHWPAEKRALYLSRVFQDSKMGTA PRMTVAENLLIARQGGKRSLASRK 122

40 Query: 142 IINYTEEFQKLIARTGNGLDRHLETPTGLSSGGQRQALSLLMATLKKPNLLLLDEHTAAL 201
I + F+ L+ RTGNGL++HLETP GLSSGGQRQALSLLMATLKKP LLLLLDEHTAAL

Sbjct: 123 ITEHLASFEDLVKRTGNGLKHELETPAGLLSGGQRQALSLLMATLKKPALLLLDEHTAAL 182

45 Query: 202 DPRTSVSLMGLTDEFIKQDSL TALMITHH MEDALKYGNRVLMKDGKIVRDLNQAQKNKM 261
DP+TS SLM LTDEF+ +D LTALMITHH MEDAL YGNR++VMKDG I++DLNQ +K ++

Sbjct: 183 DPKTGSLSMQLTDEFVTKDGLTALMITHH MEDALTYGNRLIVMKDGNIIKDLNQMEKEQL 242

Query: 262 AIADYYQLFD 271
I DYYQLFD

50 Sbjct: 243 TITDYYQLFD 252

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 895

55 A DNA sequence (GBSx0949) was identified in *S.galactiae* <SEQ ID 2721> which encodes the amino acid sequence <SEQ ID 2722>. Analysis of this protein sequence reveals the following:

Possible site: 33
>>> Seems to have no N-terminal signal sequence

60 ----- Final Results -----

-989-

bacterial cytoplasm --- Certainty=0.1930(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

5 RGD motif: 415-417

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB06117 GB:AP001515 unknown conserved protein [Bacillus halodurans]
 Identities = 236/549 (42%), Positives = 362/549 (64%), Gaps = 2/549 (0%)

10 Query: 4 IKIMALGGVRENGKNLYVVEVNDISFVLDAGLKYPENEQLGVDVVIPNLDYLIENKKRVQ 63
 I++ ALGGV E GKN+YVVEV+D +FV+DAGL +P++E LGVDVVIP++ YL+EN++RV+
 Sbjct: 9 IRVFPALGGVGEIGKNMYVVEVDDDLFVIDAGLMFPDDEMLGVDVVIPDISYLVENEERVR 68

15 Query: 64 GIFLTHGHADAIGALPYIIAEVKAPVFGSPLTIELAKLFVKNSTAVKKFNNFHVVIDSETE 123
 I LTHGH D IG LPY++ ++ PV+G+ LT+ L + +K + ++ +IDS +
 Sbjct: 69 AILLTHGHEDHIGGLPYVLQKLNVPVYGTKLTGLVEEKLKEAGLIRSAK-LKLIDSNSR 127

20 Query: 124 IEFQDAVISFFKTTTHSIPESMGIVIGTKEGNIVYTGDFKFDQAARKYYQTDLARLAEIGR 183
 ++ +SFF+T HSIPT+S+GI I T +G IV+TGDFKFDQ Q ++ ++A IG
 Sbjct: 128 LKLGSTPVSFFRTNHSIPDSVIGICIQTSQGFIVHTGDFKFDQTFVDGKQAEIGKMAAIGH 187

25 Query: 184 DGVLLALLSDSANATSNEQVASEYEVGDEIKSVIEDAEGRVIVAASNLIRIQQVFDAAA 243
 GVL LLSDS NA SE EVG I E +GR+IV ASN+ R+QQV AA
 Sbjct: 188 KGVLCLLSDSTNAERPGMTKSETEVGRGIAEAFEQTKGRIIVITTFASNVHRVQVVIHAAI 247

30 Query: 244 ENGRRVVLTGFDIENIVRTAIRMKRIHIADENMIKPKDMTRYEDNELLILETGMRGEBPI 303
 R++ + G + +V A R+ + D+ + I +++++Y+D + I+ TG GEP+
 Sbjct: 248 ATNRKLAVAGRSMVKVVSIAERLGYLEAPDD-LFIDIEVSKYDDERVAIITTGSGQBPM 306

35 Query: 304 NGLQKMAIGRHRVYQIKDGLVFIIVTTPSIAKEAVVARVENLIYKAGGSVKLITQNLRV 363
 + L +MA G HR + I + D V I TP E V+ + +L+++ G V + S
 Sbjct: 307 SALSMAKGAHRQITITENDTVIIAATPIPCNERSVSTIVDLLHRIGADVIFGHCKVHAS 366

40 Query: 364 GHANGRELQLLMNLKPKYLFPIQGEYRDLSEAGLAQEVGMSADDIYIVKRGDIMVLEK 423
 GH + EL+L++NL++PK+ PI GE+R AH LA+ VG+ + I++V +G+++
 Sbjct: 367 GHGSAEELKMLNLMRPKFFVPIHGEBFRMQAHKELAKSVGIREBAIFLVDKGEVVEFRN 426

45 Query: 424 DGFPHSGSVFAGDVMIDGNAIGDVGNIPLRDRKVLSEDFIFIVITVSKKEKKIISKARV 483
 +G VP+G+V+IDG +GDVGNIPLRDR++LS+DGI +VV+T++K+ I+S +
 Sbjct: 427 GQGRKAGKVPNGVNLIDGLGVGDVGNIPLRDRRLSKDGLVNVVVTLNKQSGTILSGPNI 486

50 Query: 484 NTRGFVYVVKSRDILRESAELVNTTVEDYLSKDTFDWGELGKGVKRVDEVSKFLFDQTKRRP 543
 +RGFVYV++S ++ E+ ELV T++ +++++ +W LK VR+ +S+FLF++TKRRP
 Sbjct: 487 ISRGFVYVRESEKLIBEANELVTETLKKCVTENVNEWSSLKSNVREVLRSFLFEKTKRRP 546

Query: 544 AILPVVMEV 552
 ILP++MEV
 Sbjct: 547 MILPIIMEV 555

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2723> which encodes the amino acid sequence <SEQ ID 2724>. Analysis of this protein sequence reveals the following:

Possible site: 33

55 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

60 bacterial cytoplasm --- Certainty=0.2204(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:BAB06117 GB:AP001515 unknown conserved protein [Bacillus halodurans]
 Identities = 232/549 (42%), Positives = 360/549 (65%), Gaps = 2/549 (0%)

-990-

Query: 4 IKMIALGGVREYGNFYLVINDSMFILDAGLKYPENEQLGVDLVIPNLDYVIENKGVQ 63
 I++ ALGGV E GKN Y+VE++D +F++DAGL +P++E LGVD+VIP++ Y++EN+ +V+
 5 Sbjct: 9 IRVFALGGVGEIGKNMYVVEVDDDLFVIDAGLMFPDDMLGVDVVIDISYLVENEERVR 68

Query: 64 GIFLSHGADAIGALPYLLAEVSAPVFGSELTIELAKLFVKSNNSTKKFNNFHVVDSDTE 123
 I L+HGH D IG LPY+L +++ PV+G++LT+ L + +K + ++DS++
 Sbjct: 69 AILLTHGHEDHIGGLPYVLQKLNVPVYGTKLTLGLVEEKLEAGLIRSAK-LKLIDNSR 127

Query: 124 IEFKDLGLVSFFRTTHSIPESMGIVIGTDKGNIIYTGDFKFDQAAREGYQTDLLRLAEIGK 183
 ++ VSFFRT HSI P+S+GI I T +G I++TGDFKFDQ +G Q ++ ++A IG
 10 Sbjct: 128 LKLGSTPVSFRTNHSIPDSVGICIQTSQGFIVHTGDFKFDQTFVDGKQARIGKMAAIGH 187

Query: 184 EGVALLSDSVNATSNQIASESEVGEEMDSVISDADGRVIVA AVASNLVRIQQVFD SAT 243
 +GVL LLSDS NA SE+EVG + GR+IV ASN+ R+QQV +A
 15 Sbjct: 188 KGVLCLLSDSTNAERPGMTKSETEVGRGIAEAFEQTGRRIIVTTFASNVRHVQQVIAHAI 247

Query: 244 AHGRRVLTGTDAENIVRTALRLEKLMITDERLLIKPKDMSKFEDHELIILEAGRMGEPI 303
 A R++ + G +V A RL L D+ L I +++SK++D + I+ G GEP+
 20 Sbjct: 248 ATNRKLAVAGRSMVKVVSIAERLGYLEAPDD-LFIDIEVSKYDDERVAIITGSGQGEPM 306

Query: 304 NSLQKMAAGRHRVYQIKEGDLVYIVTTPSTAKEAMVARVENLIYKAGGSVKLITQNL RVS 363
 ++L +MA G HR + I E D V I TP E V+ + +L+++ G V + S
 25 Sbjct: 307 SALSRLAKGAHRQITITENDTVIIAATPIGNERSVSTIVDLLHRIGADVIFGHEGKVAS 366

Query: 364 GHANGRLQLLMNLLKPQYLFVPQGEYRDLAAHAKLAEVGIFFENIHLKRGDIMVLND 423
 GH + +L+L++NL++P++ P+ GE+R AH +LA+ VGI E I ++ +G+++ +
 Sbjct: 367 GHGSAEELKLMNLMRPKFFVPIHGEFRMQHAHKLAKSVGIREAIFLVDKGEVVEFRN 426

Query: 424 EGFLEHGGVPASDV MIDGNAIGDVGNIVLRDRKVLSE DGIFIVAITVSKKEKRIISKAV 483
 G VP+ +V+IDG +GDVGNIVLRDR++LS+DGI +V +T++K+ I+S +
 30 Sbjct: 427 GQGRKAGKVPSGNVLIDGLGVGDVGNIVLRDRRLLSKDGILVVVTLNKQSGTILSGPNI 486

Query: 484 NTRGFVYVKKSHDILRESAELVNTTVGNYLKKDTPDWGELKGNVRDDLKFLFEQTKRRP 543
 +RGFVV++S ++ E+ ELV T+ + ++ +W LK NVR+ LS+FLFE+TKRRP
 35 Sbjct: 487 ISRGFVYVRESEKLEIEANELVTETLKKCVTENNVNEWSSLKSNVREVLRSFLFEKTKRRP 546

Query: 544 AILPVVMEV 552
 ILP++MEV
 40 Sbjct: 547 MILPIIMEV 555

An alignment of the GAS and GBS proteins is shown below.

Identities = 446/553 (80%), Positives = 513/553 (92%)

45 Query: 1 MSDIKMALGGVRENGKNLYVVEVNDISIFVLDAGLKYPENEQLGVDVVIPNLDYLIENKK 60
 M+DIK++ALGGVRE GKN Y+VE+NDS+F+LDAGLKYPENEQLGVD+VIPNLDY+IENK
 Sbjct: 1 MTDIKMALGGVREYGNFYLVINDSMFILDAGLKYPENEQLGVDLVIPNLDYVIENK 60

50 Query: 61 RVQGIFLTHGHADAIGALPYIIAEVKAPVFGSPLTIELAKLFVKNSTAVKKFNNFHVIDS 120
 +VQGIFL+HGHADAIGALPY++AEV APVFGS LTIELAKLFVK++ + KKFNNFHV+DS
 Sbjct: 61 KVQGIFLSHGADAIGALPYLLAEVSAPVFGSELTIELAKLFVKSNNSTKKFNNFHVVD 120

55 Query: 121 ETEIEFQDAVISFFKTTHSIPESMGIVIGTKEGNIVYTGDFKFDQAARKYYQTDLARLAE 180
 +TEIEF+D ++SFF+TTHSIPESMGIVIGT +GNI+YTGDFKFDQAAR+ YQTDL RLAE
 Sbjct: 121 DTEIEFKDGLVSFFRTTHSIPESMGIVIGTDKGNIIYTGDFKFDQAAREGYQTDLLRLAE 180

60 Query: 181 IGRDGLVALLSDSANATSNQVASEVEVGDEIKSVIEDAERGVIVA AVASNLIRIQQVFD 240
 IG++GVLALLSDS NATSN+Q+ASE EVG+E+ SVI DA+GRVIVA AVASNL+RIQQVFD
 Sbjct: 181 IGKEGVLALLSDSVNATSNQIASESEVGEEMDSVISDADGRVIVA AVASNLVRIQQVFD 240

Query: 241 AAAENGRRVLTGTDAENIVRTAIRMKRIHIADENMIKPKDMTRYEDNELLILETGRMG 300
 +A +GRRVLTG D ENIVRTA+R++++ I DE ++IKPKDM+++ED+EL+ILE GRMG
 Sbjct: 241 SATAHGRRVLTGTDAENIVRTALRLEKLMITDERLLIKPKDMSKFEDHELIILEAGRMG 300

65 Query: 301 EPINGLQKMAIGRHRVYQIKDGLVFIIVTTPSIAKEAVVARVENLIYKAGGSVKLITQNL 360
 EPIN LQKMA GRHRVYQIK+GDLV+IVTTPS AKEA+VARVENLIYKAGGSVKLITQNL
 Sbjct: 301 EPINSLQKMAAGRHRVYQIKEGDLVYIVTTPSTAKEAMVARVENLIYKAGGSVKLITQNL 360

-991-

Query: 361 RVSGHANGRELQLLMNLLKPKYLFPIQGEYRDLAHLAAGLAQEVGMSADDIYIVKRGDIMV 420
 RVSGHANGR+LQLLMNLLKP+YLFP+QGEYRDL+AHA LA+EVG+ ++I+I+KRGDIMV
 Sbjct: 361 RVSGHANGRDLQLLMNLLKPKYLFPIQGEYRDLAAHAKLAEEVGIFPENIHILKRGDIMV 420
 5
 Query: 421 LEKDGFHSGSVPAQDVMIDGNAIGDVGNIPLRDRKVLSEGGIFIVVITVSKKEKKIISK 480
 L +GF H G VPA DVMIDGNAIGDVGNIPLRDRKVLSEGGIFIV ITVSKKEK+IISK
 Sbjct: 421 LNDEGFLHEGGVPASDVMIDGNAIGDVGNIPLRDRKVLSEGGIFIVAITVSKKEKKIISK 480
 10
 Query: 481 ARVNRGTFVYVKKSRDILRESAELVNTTVEDYLSKDTFDWGELKGVKVRDEVSKFLFDQTK 540
 A+VNRGTFVYVKKSRDILRESAELVNTT +YL KDTFDWGELKG VRD++SKFLF+QTK
 Sbjct: 481 AKVNRGTFVYVKKSHDILRESAELVNTTVGNLYLKKDTFDWGELKGNVRDDLKFLFEQTK 540
 15
 Query: 541 RRPAILPVVMEVR 553
 RRPAILPVVMEVR
 Sbjct: 541 RRPAILPVVMEVR 553

There is also homology to SEQ ID 4910.

SEQ ID 2722 (GBS295) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell
 20 extract is shown in Figure 48 (lane 2; MW 89.4kDa). It was also expressed in *E.coli* as a His-fusion
 product. SDS-PAGE analysis of total cell extract is shown in Figure 167 (lane 9 & 11; MW 79kDa –
 thioredoxin fusion) and in Figure 238 (lane 3; MW 79kDa – thioredoxin fusion).

Purified Thio-GBS295-His is shown in Figure 244, lane 3.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 25 vaccines or diagnostics.

Example 896

A DNA sequence (GBSx0950) was identified in *S.agalactiae* <SEQ ID 2725> which encodes the amino
 acid sequence <SEQ ID 2726>. This protein is predicted to be tributyrin esterase. Analysis of this protein
 sequence reveals the following:

30 Possible site: 22
 >>> Seems to have a cleavable N-term signal seq.
 ----- Final Results -----
 35 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9885> which encodes amino acid sequence <SEQ ID 9886>
 was also identified.

40 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF62859 GB:AF157484 tributyrin esterase [Lactococcus lactis
 subsp. lactis]
 Identities = 154/262 (58%), Positives = 188/262 (70%), Gaps = 4/262 (1%)
 45
 Query: 21 MAFFNIEYHYSKVLGTERQVNVIPDAFEMSDDKIDDCDIPVLYLLHGMGNENSWQKRTN 80
 MA NIEY+S+VLG R+VNVIP++ ++ D DIPVLYLLHGM GNENSW R+
 Sbjct: 1 MAVINIEYYSEVLGMNRKVNVIPESSKVED--FTQTDIPVLYLLHGMGNENSWIIRSG 58
 50
 Query: 81 IERLLRHTNLIVMPSTDLAWYNTTKYGLDYFDAIAIELPKVKRFFPNMSDKREKNFIA 140
 IERL+RHTNL +VMPSTDL +Y NT YG++YFDAIA ELPKV+ FFPN+S KREKNFIA
 Sbjct: 59 IERLIRHTNLAIVMPSTDLGFVNTTYGMNYFDAIAHELKVINNFPPNLSTKREKNFIA 118
 Query: 141 GLSMGGYGAYKIALLTNRFSHAASLSGALSFDLFLFNNGNNNININYSWGFGLNNTDNI 200
 GLSMGGYGAY++AL T+ FS+AASLSG L+FD + N N YW GIFG+

-992-

Sbjct: 119 GLSMGGYGAYRLALGTDYFSYAASLSGVLTFDG--MEENFKENPAYWGGIFGNWETFKGS 176

Query: 201 ERHSLRRYVESFDMKTKFYAWCGYEDFLFEANEVAIDELRQLGLTIDYFNDHGKHEWYYW 260
+ L + K K YAWCG +DFLF NE A EL++LG I Y + G HEWYYW

5 Sbjct: 177 DNEILSLADRKQENKPKLYAWCGKQDFLPGNEYATAELKKLGFDTITYESSDGVHEWYYW 236

Query: 261 NQOLEKVLEWLPVDYVKEERLS 282
Q++E VL+WLP++Y +EERLS

10 Sbjct: 237 TQKIESVLKWLPIYKQERLS 258

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2727> which encodes the amino acid sequence <SEQ ID 2728>. Analysis of this protein sequence reveals the following:

Possible site: 14
>>> Seems to have no N-terminal signal sequence

15 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.2183(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
20 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 172/262 (65%), Positives = 199/262 (75%), Gaps = 1/262 (0%)

25 Query: 21 MAFFNIEYHSHKVLGTERQVNVIPDAFEMSDDKIDDCDIPVLYLLHGMGNGNENSWQKRTN 80
MA IEVHS VLG ER+VNVIPD E+ D DIPVLYLLHGMGNGNENSWQKRT
Sbjct: 1 MASIAIEYHSHSVLGMERKVNVIYPDQSEIPKKDQGDKDIPVLYLLHGMGNGNENSWQKRTA 60

Query: 81 IERLLRHTNLIVMPSTDLAWYTNTRYGLDYFDAIAIELPKVLKRFFPNMSDKREKNFIA 140
IERLLRHTNLIVMPSTDL WYT+T YGL+Y+ A++ ELP+VL FFPNM+ KREK F+A
30 Sbjct: 61 IERLLRHTNLIVMPSTDLGWYTDYGLNYRALSQELPQVLAFFPNMTQKREKTFVA 120

Query: 141 GLSMGGYGAYKIALLTNRFSHAASLSGALSFDLFLFNNNGNNINYNWSGIFGDLNNTDNI 200
GLSMGGYGGA+K AL +NRFS+AA SGA L F + L + YW G+FG ++ D +
35 Sbjct: 121 GLSMGGYGAFKWLKSNRFSYAASFSGALDFSPETLLEGKLGELAYWQGVFGQFDDPD-L 179

Query: 201 ERHSLRRYVESFDMKTKFYAWCGYEDFLFEANEVAIDELRQLGLTIDYFNDHGKHEWYYW 260
++H L+ V D KTKFYAWCGYEDFLF NE AI + + GL IDY HGKHEWYYW
Sbjct: 180 DKHYLKNMVAESDGKTKFYAWCGYEDFLFATNEKAIADFQAQGLDIDYHKGHGKHEWYYW 239

40 Query: 261 NQOLEKVLEWLPVDYVKEERLS 282
NQOLE +LEWLP++Y KEERLS
Sbjct: 240 NQOLEVLEWLPINYQKEERLS 261

SEQ ID 2726 (GBS645) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell
45 extract is shown in Figure 129 (lanes 8 & 10; MW 60kDa + lane 9; MW 27kDa) and in Figure 186 (lane 4;
MW 60kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell
extract is shown in Figure 129 (lane 12; MW 34.7kDa), in Figure 140 (lane 8; MW 35kDa) and in Figure
178 (lane 4; MW 35kDa). Purified GBS645-GST is shown in Figure 236, lane 11; purified GBS645-His is
shown in Figure 229, lanes 3-4.

50 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
vaccines or diagnostics.

Example 897

A DNA sequence (GBSx0951) was identified in *S.agalactiae* <SEQ ID 2729> which encodes the amino
acid sequence <SEQ ID 2730>. Analysis of this protein sequence reveals the following:

55 Possible site: 44
>>> Seems to have no N-terminal signal sequence

-993-

INTEGRAL Likelihood = -9.34 Transmembrane 22 - 38 (18 - 46)

----- Final Results -----

5 bacterial membrane --- Certainty=0.4736(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2731> which encodes the amino acid
10 sequence <SEQ ID 2732>. Analysis of this protein sequence reveals the following:

Possible site: 52

>>> Seems to have an uncleavable N-term signal seq

15 INTEGRAL Likelihood = -7.43 Transmembrane 25 - 41 (20 - 46)
INTEGRAL Likelihood = -2.71 Transmembrane 4 - 20 (3 - 20)

----- Final Results -----

20 bacterial membrane --- Certainty=0.3972(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below.

25 Identities = 31/87 (35%), Positives = 50/87 (56%), Gaps = 2/87 (2%)

Query: 1 MRTLFMRMIFAIPKFIFRLIWNIIWGIFKTVLVIAIILFGLYYYANHSQSEFANQLSDIIQ 60

M+ L +I +PK I ++ W++I G +T+L++ II+ GL YY+NHS S AN++S I

Sbjct: 1 MKQLLAIIILWLPKLIVKMFVHLIKGFLQTILLVTIIIIIGLMYYSNHSDSVLANKIS--IV 58

30 Query: 61 TGKTFILNFADTNQLKNSFTNLATDNVH 87

T + F Q ++ T + N H

Sbjct: 59 TEQVVQIFDILTQKPSAKTRHGSGNSH 85

35 SEQ ID 2730 (GBS220d) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 155 (lane 11-13; MW 50kDa) and in Figure 239 (lane 12; MW 50kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 155 (lane 14-16; MW 25.2kDa) and in Figure 184 (lane 7; MW 25kDa). Purified GBS220d-GST is shown in Figure 246, lanes 3 & 4.

40 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 898

A DNA sequence (GBSx0953) was identified in *S.agalactiae* <SEQ ID 2733> which encodes the amino acid sequence <SEQ ID 2734>. This protein is predicted to be unnamed protein product (rpiA). Analysis of this protein sequence reveals the following:

45 Possible site: 33

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

50 bacterial cytoplasm --- Certainty=0.2538(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

-994-

>GP:CAB69583 GB:A93589 unnamed protein product [Spinacia oleracea]
Identities = 114/232 (49%), Positives = 147/232 (63%), Gaps = 11/232 (4%)

5 Query: 2 DELKKLAGVTAAKYVKNGMIVGLGTGSTAYFFVVEIGRRVKEEGL-QVVGVTTSNRTTEQ 60
D+LKKLA A VK+GM++GLGTGSTA F V IG + L +VG+ TS RT EQ
Sbjct: 59 DDLKKLAAEKAVDSVKSGMVLGLGTGSTAFAVSRIGELLSAGKLTNIVGIPTSKRTAEQ 118

10 Query: 61 ARGLGIPLKSADDIDVIDVTVDGADEVDPDFNGIKGGGALLMEKIVATPTKEYIWWVDE 120
A LGIPL DD ID+ +DGADEVDPD N +KG GGALL EK+V + ++I VVD+
Sbjct: 119 AASLGIPLSVLDDHPRIDLDAIDGADEVDPDLNLVKGGRGALLREKMVEAASDKFIVVDD 178

15 Query: 121 SKLVETLGAFKL--PVEVV---RYGSERLFRVFKSKGYCPSFRETEGDR--FITDMGNY 172
+KL+ LG +L PVEVV +Y +RL +FK G C + EGD ++TD NY
Sbjct: 179 TKLVDGLGGSRLAMPVEVVQFCWKYNLKRQLQEIFKELG-CEAKLRMEGDSPPYVTDNSNY 237

Query: 173 IIDLDL-KKIEDPKQLANELDHTVGVEHGLFNGMVNKKVIVAGKNGLDILEK 223
I+DL I+D + E+ GVVEHGLF GM ++VI+AGK G+ + K
Sbjct: 238 IVDLYFPTSIIKDAEAGREISALEGVVEHGLFLGMASEVIIAGKTGVSVKTK 289

20 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2735> which encodes the amino acid sequence <SEQ ID 2736>. Analysis of this protein sequence reveals the following:

Possible site: 30
>>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.1646(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

30 An alignment of the GAS and GBS proteins is shown below.

Identities = 166/222 (74%), Positives = 190/222 (84%)

35 Query: 1 MDELKKLAGVTAAKYVKNGMIVGLGTGSTAYFFVVEIGRRVKEEGLQVVGVTTSNRTTEQ 60
M+ LKK+AGVTAA+YV +GM +GLGTGSTAY+FVEEIGRRVK+EGLQVVGVTTS+ T++Q
Sbjct: 1 MEALKKIAGVTAAQYVTDGMTIGLGTGSTAYFFVVEIGRRVKQEGLQVVGVTSSVTSKQ 60

40 Query: 61 ARGLGIPLKSADDIDVIDVTVDGADEVDPDFNGIKGGGALLMEKIVATPTKEYIWWVDE 120
A LGIPLKS DDID ID+TVDGADEVDPDFNGIKGGG ALLMEKIVATPTKEYIWWVD
Sbjct: 61 AEVLGIPLKSIDDIDSIDLTVDGADEVDPDFNGIKGGGALLMEKIVATPTKEYIWWVDA 120

Query: 121 SKLVETLGAFKLPEVVRYGSERLFRVFKSKGYCPSFRETEGDRFITDMGNYIIDLDLKK 180
SK+VE LGAFKLPEVV+YG++RLFRVF+ GY PSFR R +TDM NYIIDLDL
Sbjct: 121 SKMVEHLGAFKLPEVVQYGADRLFRVFEKAGYKPSFRMKGDSRLVTDMQNYIIDLDLGC 180

45 Query: 181 IEDPKQLANELDHTVGVEHGLFNGMVNKKVIVAGKNGLDILE 222
I+DP + LD TVGVVEHGLFNGMV+KVIVA K+G+ +LE
Sbjct: 181 IKDPVAFGHLLDGTGVVEHGLFNGMVDKVIVASKDGVTVLE 222

50 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 899

A DNA sequence (GBSx0954) was identified in *S.galactiae* <SEQ ID 2737> which encodes the amino acid sequence <SEQ ID 2738>. This protein is predicted to be phosphopentomutase (deoB). Analysis of this protein sequence reveals the following:

55 Possible site: 22
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.0546(Affirmative) < succ>

-995-

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

5 >GP:AAC45496 GB:U80410 phosphopentomutase [Lactococcus lactis subsp. cremoris]
 Identities = 275/408 (67%), Positives = 325/408 (79%), Gaps = 7/408 (1%)

Query: 3 QFDRIHLVVLDSVGIGAAPDANDFVNAGVP-----DCASDTLGHISKT VGLAVPNMAKI 56
 +F RIHLVV+DSVGIGAAPDA+ F N V D SDT+GHIS+ GL VPN+ K+

10 Sbjct: 4 KFGRIHLVVMDSVGIGAAPDADKFFNHVDVTHEAINDVKSDTIGHISEIRGLDVPNLQKL 63

Query: 57 GLGNIPRPQALKTVPAEENPSGYATKLQEVSLGKDTMTGHWEIMGLNITEPFDTFWNGFP 116
 G GNIPR LKT+PA + P+ Y TKL+E+S GKDTMTGHWEIMGLNI PF T+ G+P

15 Sbjct: 64 GWGNIPRESPLKTI PAAQKPAAYVTKLEBISKGKDTMTGHWEIMGLNIQT'PFPTYPEGY 123

Query: 117 EDIITKIEDFSGRKVI REANKPYSGTAVIDDFGPRQMETGELI IYTSADPVLQIAAHEDI 176
 ED++ KIE+FSGRK+IREANKPYSGTAVI+DFGPRQ+ETGELI IYTSADPVLQIAAHED+

Sbjct: 124 EDLLEKIEEFSGRK IIREANKPYSGTAVIEDFGPRQLETGELI IYTSADPVLQIAAHEDV 183

20 Query: 177 IPLEELYRICEYARSITMERFALL-GRIIARPYVGEPGNFTRTANRHDYAVSPFEDTVLN 235
 I EELY+ICEY RSIT+E ++ GRIIARPYVGE GNF RT R DYA+SPF +TVL

Sbjct: 184 ISREELYKICEYVRSITLEGGIMIGRIIARPYVGEAGNFERTDGRDYALSPFAETVLE 243

25 Query: 236 KLDQAGIDITYAVGKINDIFNGSGINHDGMHNKSNHSHGIDTLIKTMGLSEFEKGFSTNLV 295
 KL +AGIDITY+VGKI+DIFN G+ +DMGHN ++ G+D L+K M +EF +GFSFTNLV

Sbjct: 244 KLYKAGIDITYSVGKISDIFNTVGKIDMGHNHNDMDGVDRLLKAMTKTEFTTEGFSFTNLV 303

Query: 296 DFDALYGHRRDPHGYRDCLEHFDERLPEIISAMRDKDLLITADHGNDPTYAGTDHTREY 355
 DFDA YGHRRD GY + +FD RLPEII AM++ DLL+ITADHGNDP+Y GTDHTREY

30 Sbjct: 304 DFDKAYGHRRDVEGYGAIEDFDGRLPEIIDAMKEDLLMITADHGNDPSYVGTDHTREY 363

Query: 356 IPLLAYSPTSFTGNGLIPVGHFADISATVADNFGVDTAMIGESFLQDLV 403
 IPL+ +S SF ++PVGHFADISAT+A+NF V A GESFL LV

35 Sbjct: 364 IPLVIFSKSFKEPKVLPVGHFADISATIAENFVSKAQTGESFLDALV 411

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2739> which encodes the amino acid sequence <SEQ ID 2740>. Analysis of this protein sequence reveals the following:

Possible site: 22
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0185(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 348/402 (86%), Positives = 374/402 (92%)

50 Query: 1 MSQFDRIHLVVLDSVGIGAAPDANDFVNAGVDPGASDTLGHISKT VGLAVPNMAKIGLGN 60
 MS+F+RIHLVVLDSVGIGAAPDA+ F NAGV D SDTLGHIS+ GL+VPNMAKIGLGN

Sbjct: 1 MSKFNRHLVVLDSVGIGAAPDADKFFNAGVADTSDTLGHISEAAGLSVPNMAKIGLGN 60

Query: 61 IPRPQALKTVPAEENPSGYATKLQEVSLGKDTMTGHWEIMGLNITEPFDTFWNGFPEDII 120
 I RP LKTVP E+NP+GY TKL+EVSLGKDTMTGHWEIMGLNITEPFDTFWNGFPE+I+

55 Sbjct: 61 ISRPIPLKTVPTEDNPTGYVTKLEEVSLGKDTMTGHWEIMGLNITEPFDTFWNGFPPEIL 120

Query: 121 TKIEDFSGRKVI REANKPYSGTAVIDDFGPRQMETGELI IYTSADPVLQIAAHEDIIPLE 180
 TKIE+FSGRK+IREANKPYSGTAVIDDFGPRQMETGELI+YTSADPVLQIAAHEDIIP+E

60 Sbjct: 121 TKIEEFSGRK IIREANKPYSGTAVIDDFGPRQMETGELIVYTSADPVLQIAAHEDIIPVE 180

Query: 181 ELYRICEYARSITMERFALLGRIIARPYVGEPGNFTRTANRHDYAVSPFEDTVLNKLDQA 240
 ELY+ICEYARSIT+ERPALLGRIIARPYVG+PGNFTRTANRHDYAVSPF+DTVLNKL A

Sbjct: 181 ELYKICEYARSITLERPALLGRIIARPYVGDPGNFTRTANRHDYAVSPFQD TVLNKLADA 240

-996-

Query: 241 GIDTYAVGKINDIFNGSGINHDMGHNKSNSHGIDTLIKTMGLSEFEKGFSTNLVDFDAL 300
 G+ TYAVGKINDIFNGSGI +DMGHNKSNSHGIDTLIKT+ L EF KGFSFTNLVDFDA
 Sbjct: 241 GVPTYAVGKINDIFNGSGITNDMGHNKSNSHGIDTLIKTLQLPEFTKGFSFTNLVDFDAN 300

5 Query: 301 YGHRDPHGYRDCLEHFDERLPEIISAMRDKDLLLITADHGNDPTYAGTDHTREYIPLLA 360
 +GHRDP GYRDCLEHFD RLPEII+ M++ DLLLITADHGNDPTYAGTDHTREYIPLLA
 Sbjct: 301 FGHRDPGYRDCLEHFDNRLPEIIANMKEDDLLLITADHGNDPTYAGTDHTREYIPLLA 360

10 Query: 361 YSPSFTGNGLIPVGHFADISATVADNFGVDTAMIGESFLQDL 402
 YS SFTGNGLIP GHFADISATVA+NFGVDTAMIGESFL L
 Sbjct: 361 YSVSFTGNGLIPQGHFADISATVAENFGVDTAMIGESFLSHL 402

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

15 Example 900

A DNA sequence (GBSx0955) was identified in *S.galactiae* <SEQ ID 2741> which encodes the amino acid sequence <SEQ ID 2742>. This protein is predicted to be unnamed protein product (mtaP). Analysis of this protein sequence reveals the following:

Possible site: 36
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -1.44 Transmembrane 215 - 231 (215 - 231)

----- Final Results -----
 bacterial membrane --- Certainty=0.1574(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2743> which encodes the amino acid sequence <SEQ ID 2744>. Analysis of this protein sequence reveals the following:

Possible site: 36
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -1.44 Transmembrane 215 - 231 (215 - 231)

----- Final Results -----
 bacterial membrane --- Certainty=0.1574(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

40 An alignment of the GAS and GBS proteins is shown below.

Identities = 225/269 (83%), Positives = 248/269 (91%)

Query: 1 MTLLEKINETRDFLQAKGVTAPEFGLILGSGLGELAEIENPIVVDYADIPNWGQSTVVG 60
 M+L+ KINET+DFL KG+ PEFGLILGSGLGELAE+EN IV+DYADIPNWG+STVVG
 45 Sbjct: 1 MSLMTKINETKDFLVTKGIETPEFGLILGSGLGELAEVENAIVIDYADIPNWGKSTVVG 60

Query: 61 HAGKLVYGDLSGRKVLALQGRFHFYEGNTMEVVTFFPVRIMRALACHSVLVTNAAGGIGYG 120
 HAGKLVYGD+GRKVLALQGRFHFYEGN +EVVTFPVR+M+AL C VLVNTAAGGIGYG
 50 Sbjct: 61 HAGKLVYGDLAGRKVLALQGRFHFYEGNPLEVVTFFPVRVMKALGCEGVLTNAAGGIGYG 120

Query: 121 PGTLMMLIKDHINMIGTNPLIGENLBEFGPRFPDMSDAYTATYRQKAHQIAEKQNIKLEEG 180
 PGTLM I DHINM G NPLIGENL+EFGRFPDMSDAYT YR KAH++AEK NIKLE+G
 55 Sbjct: 121 PGTLMAITDHINMTGNNPLIGENLDEFGRFPDMSDAYTKVYRNKAHEVAEKMNIKLEEDG 180

Query: 181 VYLGVSQPTTYETPAEIRAFQTMGAQAVGMSTVPEVIVAAHSGLKVLGISAITNFAAGFQS 240
 VY+G++GPTYETPAEIRAF+ +GA AVGMSTVPEVIVAAHSGLKVLGISAITNFAAGFQS
 60 Sbjct: 181 VYMGLTGPTYETPAEIRAFKVLGADAVGMSTVPEVIVAAHSGLKVLGISAITNFAAGFQS 240

Query: 241 ELNHEEVVEVTQRIKEDFKGLVKLSLVAEL 269

-997-

ELNHEEVVEVTQ IKEDFKGLVK+++AEL
 Sbjct: 241 ELNHEEVVEVTQHIKEDFKGLVKAILAEL 269

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 5 vaccines or diagnostics.

Example 901

A DNA sequence (GBSx0956) was identified in *S.galactiae* <SEQ ID 2745> which encodes the amino acid sequence <SEQ ID 2746>. Analysis of this protein sequence reveals the following:

Possible site: 31
 10 >>> Seems to have a cleavable N-term signal seq.
 INTEGRAL Likelihood = -9.34 Transmembrane 266 - 282 (263 - 289)
 INTEGRAL Likelihood = -8.97 Transmembrane 231 - 247 (229 - 253)
 INTEGRAL Likelihood = -7.70 Transmembrane 356 - 372 (352 - 376)
 15 INTEGRAL Likelihood = -7.32 Transmembrane 303 - 319 (297 - 326)
 INTEGRAL Likelihood = -5.57 Transmembrane 337 - 353 (334 - 355)
 INTEGRAL Likelihood = -5.57 Transmembrane 391 - 407 (387 - 409)
 INTEGRAL Likelihood = -2.44 Transmembrane 177 - 193 (177 - 193)
 INTEGRAL Likelihood = -1.01 Transmembrane 159 - 175 (159 - 175)
 20 INTEGRAL Likelihood = -0.43 Transmembrane 198 - 214 (196 - 215)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.4736(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 25 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9883> which encodes amino acid sequence <SEQ ID 9884> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

30 >GP:AAD53928 GB:AF179611 chloride channel protein [Zymomonas
 mobilis]
 Identities = 121/410 (29%), Positives = 213/410 (51%), Gaps = 19/410 (4%)
 Query: 14 VKFMIAVLFMTVMAGVGAILMHYVLMFTFWLAFGDSRENTLSLLN-----SVTPIKRVL 67
 +++ +A L + + G+G +L+ ++L + +A+G S ++ +S + + +P++R+
 35 Sbjct: 3 IRYGLACLA VGC LTGLGGM LLSWILHAVQHIA YGYS LQHVISEESFLKGSMAASPLRRLE 62
 Query: 68 SLTLVSFLASLSWYWLQIKPKQITSIKQQVVF KD FSVK KSPYWLHIGHAFLQLIYVGTGG 127
 L + W L+ + SI Q V + P+W I H LQ++ VG G
 40 Sbjct: 63 VLVFCGAVVGGGWGLLRHFGSPLVSITQAVAANK---RVMPFWTTI IHVLLQIVTVGLGS 119
 Query: 128 PIGKEGAPREFGAINAGKISDLLALKVLDKRLLIISGAAAGLSAVYQVPLASVFFAFETL 187
 P+G+E APRE G++ + + L +R+L+ GA AG ++VY VPL+ FA E L
 Sbjct: 120 PLGREVAPRELGS LIGERFAFWGGLSENQRRLVACGAGAGFASVYNVPLSGALFALEAL 179
 45 Query: 188 ALGISLKNIVTLLASTFGAASIAQLVISTAPLYHISKMSLSQSLAFMFLIVLCVTPI-- 245
 + + ++ L ++ +A +A +++ + +YH+ +++++ + L+ L PI
 Sbjct: 180 LMTWASPVVIVALLTSALSARMAWILLGNSMVYHVPAPVPDTR---LMLLALLAGPIFG 235
 Query: 246 --AISFRYLNQKVTERRIK-NIKILLSLPVSVLSIVYPQILGNGNALVQEVFKGT 302
 A FR+ +QK+T RIK N ++ L + + +LS+ +P+ILNG V F
 50 Sbjct: 236 IAAHYFRFWSQKITASRIKDNRRLLVAILCFAAIGLLSMWFP EILGNGKGPVSLAFNDN 295
 Query: 303 TVSLIA-ILVVLKMIATLSITYAGAYGGILTSPFSIGACLGFLASISIPLLPHISIVTS 361
 + A L K++A L+AGAYGG+LTP S GA L ++ + LP + I
 55 Sbjct: 296 LSGMKAGELFCFKILAVFLALWAGAYGGLLTPGISFGALLAVVIGHLWNMMLPPVPIGAF 355
 Query: 362 MLVGAAIFLAITMRAPLTAVGLVISFTGQSVITIVPLTIAVLFAATAYDYF 411
 ++G A FLA +M+ P+TA+ LVI F ++P+ AV + A F
 60 Sbjct: 356 AIIGGAFLASSMKMPITAMALVIEFARTGHDFLIPIAFVAGSIAISQF 405

-998-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2747> which encodes the amino acid sequence <SEQ ID 2748>. Analysis of this protein sequence reveals the following:

```

Possible site: 13
>>> Seems to have no N-terminal signal sequence
5   INTEGRAL    Likelihood = -5.41   Transmembrane  247 - 263 ( 245 - 267)
    INTEGRAL    Likelihood = -5.15   Transmembrane  326 - 342 ( 323 - 345)
    INTEGRAL    Likelihood = -5.04   Transmembrane  411 - 427 ( 407 - 429)
    INTEGRAL    Likelihood = -4.94   Transmembrane   39 - 55 ( 34 - 59)
    INTEGRAL    Likelihood = -4.46   Transmembrane  284 - 300 ( 282 - 307)
10  INTEGRAL    Likelihood = -3.45   Transmembrane  380 - 396 ( 376 - 400)
    INTEGRAL    Likelihood = -2.13   Transmembrane  185 - 201 ( 184 - 201)
    INTEGRAL    Likelihood = -2.02   Transmembrane   88 - 104 ( 87 - 105)
    INTEGRAL    Likelihood = -1.12   Transmembrane  350 - 366 ( 350 - 367)

15  ----- Final Results -----
        bacterial membrane --- Certainty=0.3166(Affirmative) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>
        bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

20  The protein has homology with the following sequences in the databases:
    >GP:AAF41386 GB:AE002449 chloride channel protein-related protein
      [Neisseria meningitidis MC58]
      Identities = 137/373 (36%), Positives = 201/373 (53%), Gaps = 23/373 (6%)

25  Query: 59  IHLIQSLSFQFSQ----SFSTMIASVPPQRRALLSLFAGLLAGLGHLLAKKGKDIQSI 114
      +H IQ  ++G+      SF  +A      RR  L  G  +AG GW LL + GK  I
      Sbjct: 1  MHFIQHTAYGYGADGVYTSFREGVAQASGMRRVAVLTLGAVAGSGWLLKRFQKQPIEI 60

30  Query: 115 QQIIQDDISFSPW-TQFWHGWLQLTTVSMGAPVGREASREVAVTLTSLWSQRCNLSKAD 173
      + ++ +  P+ T  +H  LQ+ TV +G+P+GRE A RE+      +R  L  + +
      Sbjct: 61 KAALKQPLQGLPFLTTTFHVLLQIITVGLGSPLGREVAPREMTAAFAFAGGKRLGLDEGE 120

      Query: 174 QKLLACASGAALGAVYNAPLATILFILEAILNRWSLKNIYAACLTSYVAVETVALLQGR 233
      +LL+ACASGA L AVYN PLA+ LFILEA+L W+ + + AA LTS +A      + G
35  Sbjct: 121 MRLLIACASGAGLAAYNVPLASTLFILEAMLGVWTTQQAVALLTSVIATAVARI--GL 178

      Query: 234 HEIQYLMPPQHWTLGT--LIGSVLAGLILSLFAHAYKHLKHLPKADAKSQWFIPKVLIA 291
      ++Q  P  + T+ T  L  S + G IL + A ++  + P      + IP  +
40  Sbjct: 179 GDVQQXHP-ANLTVNTSLWFSVAVIGPILGVAVFFQRTAQKFPFIKRDNIKIPLAVCM 237

      Query: 292 FSLIAGLSIFFPEILGNKGAG--LLF-FLHEEPH---LSYISWLLVAKAVAI SLVFASGA 345
      F+LI  +S++FPEILGNKGAG L F L + H  L+ + WL+V A+A+      GA
      Sbjct: 238 FALIGVISVWFPEILGNKGAGNQLTFGGLTDWQHSGLTAVKWLVLVLMALAV-----GA 291

45  Query: 346 KGGKIAPSMMLGCGASGLLLAILLSQYLIPLSLSLNTLAIMVGATIFLGVINKIPLAAPVFLV 405
      GG I PSMMLG      A      + P  +S+ A +VGA +FLGV K+PL A F++
      Sbjct: 292 YGGLITPSMMLGSTIAFAAATAWNSVFP-EMSSESAIVGAAVFLGVSLKMPLTAIAFIL 350

      Query: 406 EITGQSLMIPL 418
      E+T      + +++PL
50  Sbjct: 351 ELTYAPVALLMPL 363

```

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 131/415 (31%), Positives = 215/415 (51%), Gaps = 9/415 (2%)

55  Query: 2  LNFKMVSRLYYAVKFMIAVLFMT-VMAGVGAILMHYVLMFTEWLAFGDSRENTLSLLNSV 60
      LNF  S +      +  LF+T + AG+ A ++  +  + L+FG S+ + +++ SV
      Sbjct: 22 LNFCYNLSMKRHFLLLTFFYLFLTGLTAGLVAFILTKAIHLIQSLSFQFSQGSFSTMIASV 81

60  Query: 61 TPIKRVLSLTLVSLASLSWYYLQIKPKQITSIKQQVVFKDFSVKKSPLYWLHIGHAFLQL 120
      P +R LSL      LA L W+ L  K K I SI QQ++ D S  SP W      H +LQL
      Sbjct: 82 PPQRRALLSLFAGLLAGLGHLLAKKGKDIQSI-QQIIQDDISF--SP-WTQFWHGWLQL 137

      Query: 121 IYVGTGGPIGKEGAPREFGAINAGKISDLLALKVLDKRLLIISCAAAGLSAVYQVPLASV 180

```


-999-

V G P+G+EGA RE S L D++LL+ + A L AVY PLA++
 Sbjct: 138 TTVSMGAPVREGASREAVTLTSLWSQRCNLSKADQKLLACASGAALGAVYNAPLATI 197
 Query: 181 FFAFETIALGISLKNIVTLLASTFGAASIAQLVISTAPL-YHISKMSLNSQSLAFMFLIV 239
 F E + SLKNI +++ A L+ + Y + + +L L
 Sbjct: 198 LFILEAILNRWSLKNITYAACLTSYVAVETVALLQGRHEIQYLMPOQHWTLGTLIGSVLAG 257
 Query: 240 LCVTPIAISFRYLNQKVTERRIKNIKILLSLPVVSLIVSVLSIVYPQILGNGNA-LVQEV 298
 L ++ A ++++L + + + K+ + + + +++ LSI +P+ILGNG A L+ +
 Sbjct: 258 LILSLFAHAYKHLKHLPKADAKSQWFIPKVLIAFSLIAGLSIFFPEILGNGKAGLLFFL 317
 Query: 299 FKGTTVSLIAILVVLKMIATLSTLYAGAYGGILTPSFSIGACLGFLASISIPLLP-HIS 357
 + +S I+ L+V K +A +GA GG + PS +G G LLA +S L+P +S
 Sbjct: 318 HEEPHLSYISWLLVAKAVAISLVFASGAKGGKIAPSMMLGGASGLLLAILSQYLIPLSLS 377
 Query: 358 IVTSMVLVGAIFLAITMRAPLTAVGLVISFTGQSVITIVPLTIA-VLFATAYDYF 411
 +++VGA IFL + + PL A ++ TGQS++ I+PL +A ++F +Y ++
 Sbjct: 378 NTLAIMVGATIFLGVINKIPLAAPVFLVEITGQSLMIPLALANLIFYFSYQFY 432

20 A related GBS gene <SEQ ID 8683> and protein <SEQ ID 8684> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 9
 SRCFLG: 0
 25 McG: Length of UR: 19
 Peak Value of UR: 2.96
 Net Charge of CR: 2
 McG: Discrim Score: 9.64
 GvH: Signal Score (-7.5): 1.15
 Possible site: 26
 30 >>> Seems to have a cleavable N-term signal seq.
 Amino Acid Composition: calculated from 27
 ALOM program count: 9 value: -9.34 threshold: 0.0
 INTEGRAL Likelihood = -9.34 Transmembrane 261 - 277 (258 - 284)
 INTEGRAL Likelihood = -8.97 Transmembrane 226 - 242 (224 - 248)
 35 INTEGRAL Likelihood = -7.70 Transmembrane 351 - 367 (347 - 371)
 INTEGRAL Likelihood = -7.32 Transmembrane 298 - 314 (292 - 321)
 INTEGRAL Likelihood = -5.57 Transmembrane 332 - 348 (329 - 350)
 INTEGRAL Likelihood = -5.57 Transmembrane 386 - 402 (382 - 404)
 INTEGRAL Likelihood = -2.44 Transmembrane 172 - 188 (172 - 188)
 40 INTEGRAL Likelihood = -1.01 Transmembrane 154 - 170 (154 - 170)
 INTEGRAL Likelihood = -0.43 Transmembrane 193 - 209 (191 - 210)
 PERIPHERAL Likelihood = 1.22 61
 modified ALOM score: 2.37
 icml HYPID: 7 CFP: 0.474
 45 *** Reasoning Step: 3
 ----- Final Results -----
 bacterial membrane --- Certainty=0.4736(Affirmative) < succ>
 50 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

ORF00327(340 - 1533 of 1869)
 55 GP|5834362|gb|AAD53928.1|AF179611_12|AF179611(3 - 405 of 425) chloride channel protein
 {Zymomonas mobilis}
 %Match = 14.7
 %Identity = 30.2 %Similarity = 56.1
 Matches = 121 Mismatches = 169 Conservative Sub.s = 104
 60
 270 300 330 360 390 420 450 468
 RSLKLLSVLKKISRD*LNH*LLNFKMVSRLYYAVKFMIAVLFMTVMAGVGAILMHYVLMFTEWLAFGDSRENTLS----L
 ::::| | : : | : | : : | : : | : : | : : |
 MKIRYGLACLAVGCLTGLGGMLLSWILHAVQHIAYGYSLOHVISEESFL
 65 10 20 30 40

-1000-

```

492      522      552      582      612      642      672      702
LNSV--TPIKRVLSLTLVSLFASLSWYYLQIKPKQITSIKQQVVFKDFSVKKSPLYWLHGHAFLLQIYVGTGGPIGKEGA
| : :|:|: | : : | : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
5  KGSMAASPLRRLEVLVFCGAVVGGWGLLRHFGSPLVSITQAVAANK--RVMPFWTTIIHVLLQIVTVGLGSPLGREVA
      60      70      80      90      100     110     120

732      762      792      822      852      882      912      942
PREFGAINAGKISDLLALKVLDKRLIIISGAAGLSAVYQVPLASVFFAFETLALGISLKNIVTLLASTFGAASIAQLVI
|||:|:|: : : | : |:|: || ||::|| |||: |:|: | : : : : | : : | : | : : :
10 PRELGSLIGERFAFWGGLSENQRRILVACGAGAGFASVYNVPLSGALFALEALLMTWASPVVIVALLTSALSARMAWILL
      140     150     160     170     180     190     200

972      1002     1032     1059     1089     1119     1146     1176
STAPLYHISKMSLNSQSLAFMFLIVLCVTPIAIS-FRYLNQKVTERRIKNIKILLSLPVVSLLI-VSVLSIVYPQILNGN
: :|:|: : : : | : | : : || ||:|:| |||: : | : : : : :|:|: :|:|:|
15 GNSMVYHVPAPVDTR-LMLLALLAGPIFGIAAHYFRFWSQKITASRIKDNRRLLVAILCFAAIGLLSMWFPETLNGK
      220     230     240     250     260     270     280

1206     1233     1263     1293     1323     1353     1383     1413
ALVQEVFKGTITVSLIA-ILVVLKMIATLSTLYAGAYGGILTPSFSIGACLGFLLASISIPLLPHISIVTSMIVGAIFLA
| | | : | | :|:|: | : |:|:|:|:| | | | : : : | : | : : | : | : | : | : | : | : | : |
20 GPVSLAFNDNLSGMKAGELFCFKILAVFLALWAGAYGGLLTPGISFGALLAVVIGHLWNMWLPPVPIGAFAIIGGAFLA
      300     310     320     330     340     350     360

1443     1473     1503     1533     1563     1593     1623     1653
ITMRAPLTAVGLVISFTGQSVITIVPLTIAVLFATAYDYFIRKMRSLYVNPY*SKTR*NCR*NFTSRRSTPCEIYCREFF
:|: |:|:|: ||| | :|: || : | |
25 SSMKMPITAMALVIEFARTGHDFLIPIAFAVAGSIAISQFYDQKKQPKTASKSVISHLGG
      380     390     400     410     420
30

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 902

35 A DNA sequence (GBSx0957) was identified in *S. agalactiae* <SEQ ID 2749> which encodes the amino acid sequence <SEQ ID 2750>. This protein is predicted to be purine nucleoside phosphorylase, fragment (deoD-1). Analysis of this protein sequence reveals the following:

```

Possible site: 25
>>> Seems to have no N-terminal signal sequence
40
----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2384 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
45      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:CAC18350 GB:Y17900 putative purine-nucleotide phosphorylase
[Streptococcus salivarius]
Identities = 200/236 (84%), Positives = 219/236 (92%)
50
Query: 1  MSIHIEAKQGEIADKILLPGDPLRAKFIAENFLEDAVCFNTVRNMFYGTGTGKGRVSV 60
        MSIHI AKQGEIADKILLPGDPLRAKFIAENFLEDAVCFN VRNMFYGTGTGKGRVSV
Sbjct: 1  MSIHIAAKQGEIADKILLPGDPLRAKFIAENFLEDAVCFNEVRNMFYGTGTGKGRVSV 60

Query: 61  GTGMGMPSISYARELIVDYGKTLIRVGTTAGAINPDIHVRELVLQAQAATNSNIIRNDW 120
        GTGMGMPSISYARELIVDYGK LIRVGTTAG++N D+HVRELVLQAQAATNSNIIRNDW
Sbjct: 61  GTGMGMPSISYARELIVDYGKTLIRVGTTAGSLNEDVHVRELVLQAQAATNSNIIRNDW 120

55
Query: 121  PEFDFPQIADFLLDKAYHIAKEMDITTHVGSVLSSDVFYSNQPDNRNMGKLGVAHAIEM 180
        P++DFPQIA+F LLDKAYHIAK +TTHVG+VLSSDVFYSN ++N+ LGK GV A+EM
60
Sbjct: 121  PQYDFPQIANFNLLDKAYHIAKNFGMTTHVGNVLSSDVFYSNYFEKNIELGKVGKAVEM 180

```

-1001-

Query: 181 EAAALYYLAAQHNVNALAMMTISDNLNPEEDTSAEERQTTFTDMMKVGLLETISE 236
 EAAALYYLAAQH V+ALA+MTISD+L NP+EDT+AEERQ TPTDMMKVGLLETI++
 Sbjet: 181 EAAALYYLAAQHVDALAIMTISDSLVNPDEDTTAEERQNTFTDMMKVGLLETIAD 236

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2751> which encodes the amino acid sequence <SEQ ID 2752>. Analysis of this protein sequence reveals the following:

Possible site: 25
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2117(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 210/235 (89%), Positives = 226/235 (95%)

Query: 1 MSIHIEAKQGEIADKILLPGDPLRAKFAENFLEDAVCFNTVRNMPGYTGTYKGHRVSVM 60
 MSIH I AK+G+IADKILLPGDPLRAKFAENFLEDAVCFN VRNMPGYTGTYKGHRVSVM
 Sbjet: 1 MSIHISAKGDIADKILLPGDPLRAKFAENFLEDAVCFNEVRNMPGYTGTYKGHRVSVM 60
 Query: 61 GTGMGMPSISIIYARELIVDYGVKTLIRVGTAGAINPDIHVRELVLQAQAATNSNIIRNDW 120
 GTGMGMPSISIIYARELIVDYGVKTLIRVGTAGAI+P++HVRELVLQAQAATNSNIIRND+
 Sbjet: 61 GTGMGMPSISIIYARELIVDYGVKTLIRVGTAGAIDPEVHVRELVLQAQAATNSNIIRNDF 120
 Query: 121 PEFDFPQIADFLLDKAYHIAKEMDITTHVGSVLSSDVFYSNPDRNMALGKLGVAIEM 180
 PEFDFPQIADF LLDKAYHIA+EM +TTHVG+VLSSDVFY+N P+RNMALGKLGVAIEM
 Sbjet: 121 PEFDFPQIADFGLLDKAYHIAREMGVTTTHVGNVLSSDVFYTNMPERNMALGKLGVAIEM 180
 Query: 181 EAAALYYLAAQHNVNALAMMTISDNLNPEEDTSAEERQTTFTDMMKVGLLETIS 235
 EAAALYYLAAQH+V AL +MTISDNLN+P EDT+AEERQTTFTDMMKVGLLETI+
 Sbjet: 181 EAAALYYLAAQHVKALGIMTISDNLNPDPTEDTTAEERQTTFTDMMKVGLLETIA 235

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 903

A DNA sequence (GBSx0958) was identified in *S.galactiae* <SEQ ID 2753> which encodes the amino acid sequence <SEQ ID 2754>. Analysis of this protein sequence reveals the following:

Possible site: 36
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1710(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9881> which encodes amino acid sequence <SEQ ID 9882> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2755> which encodes the amino acid sequence <SEQ ID 2756>. Analysis of this protein sequence reveals the following:

Possible site: 21
 >>> Seems to have no N-terminal signal sequence

-1002-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1386(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

5

An alignment of the GAS and GBS proteins is shown below.

Identities = 126/253 (49%), Positives = 175/253 (68%), Gaps = 2/253 (0%)

Query: 3 IEMTDFSTALKVLVDQYSYHNAFLLLQKHGPLNSDLLFLEMMKERRELNIDFLFAHQEQ 62
 + MT+ T L +L+D Y+Y++AF + + + L+LLEM+KERRELN+ FL H +
 Sbjct: 1 LPMTNQF-LDILLDVYAYNHAFAKALPNIPKTALYLLEMLKERRELNLAFLAEHAAE 59

Query: 63 VVILQEKYNIKL-LHNPYDELLANYIMDLKAKVKNGLIIDFVRSVSPILYRLFMILLAQ 121
 ++++Y+ L L+ + E +ANYI+DLE KVKNG IIDFVRSVSPILYRLF+ L+
 Sbjct: 60 NRTIEDQYHCSLWLNQSLDEQIANYILDLEVKVKNGLIIDFVRSVSPILYRLFRLITS 119

Query: 122 EVPHLHDYIHNARDHDYDTWKFKELKESNHPVLLAFSERWHSRLTSKSLAECLQLTDLD 181
 E+P+ YI + ++D YDTW F+ + ES+H V A+ + +T+KSLA+ L LT L
 Sbjct: 120 EIPNFKAYIFDTKNDQYDTWHFQAMLESDEHVFKAQYLSQKQSRNVITKSLADMLTLTSLP 179

Query: 182 EEVKSTIIQLRQFEKSVRNPLAHLIKPFDEQEYRTTQFSSQAFLDQIIFLAKVIGVEYD 241
 +E+K + LR FEK+VRNPLAHLIKPFDE+EL+RTT FSSQAFL+ II LA GV Y
 Sbjct: 180 QEIKDLVFLLRHFEKAVRNPLAHLIKPFDEEELHRTTHFSSQAFLNITITATFSGVIYR 239

Query: 242 TVNFHYDTVNKLI 254
 F++D +N +I
 Sbjct: 240 REPFYFDDMNIAI 252

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 904

A DNA sequence (GBSx0959) was identified in *S.agalactiae* <SEQ ID 2757> which encodes the amino acid sequence <SEQ ID 2758>. This protein is predicted to be CpsY protein. Analysis of this protein sequence reveals the following:

Possible site: 35
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.59 Transmembrane 260 - 276 (260 - 276)

----- Final Results -----
 bacterial membrane --- Certainty=0.1235(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9879> which encodes amino acid sequence <SEQ ID 9880> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2759> which encodes the amino acid sequence <SEQ ID 2760>. Analysis of this protein sequence reveals the following:

Possible site: 35
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1958(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

-1003-

Identities = 247/301 (82%), Positives = 274/301 (90%)

Query: 1 MRIQQLQYVIKIVETGSMNEAAKQLYITQPSSLNAVNRNLETEMGIQIFIRNPKGITLTKD 60
 MRIQQL Y+IKIVE GSMNEAAKQL+ITQPSSLNAV++LE EMGI IF RNPKGITLTKD
 5 Sbjet: 1 MRIQQLHYIIKIVECGSMNEAAKQLFITQPSSLNAVKDLEMEMGITIFNRNPKGITLTKD 60

Query: 61 GMEFLSYARQILEQTALLEERYKGDNTSRELFVSSQHYAFVVNAFVAFNGTDMTQYEL 120
 G+EFLSYARQI+EQT+LLE+RYK NT RELFSVSSQHYAFVVNAFV+L TDMT+YEL
 10 Sbjet: 61 GVEFLSYARQIIEQTSLLEDYKHNHTGRELFSVSSQHYAFVVNAFVSLKRTDMTRYEL 120

Query: 121 FLRETRTWEIIDDVKNFRSEIGVLFINSYNRDVLTCLFDDNSLIATTLFTTTPHIFVSKS 180
 FLRETRTWEIIDDVKNFRSEIGVLF+N YNRDVLTKLFDDN L A+ LF PHIFVSKS
 15 Sbjet: 121 FLRETRTWEIIDDVKNFRSEIGVLFINDYNRDVLTCLFDDNHLTASPLFKAQPHIFVSKS 180

Query: 181 NPLANRKKLNMKDLEDYPYLSYDQGLHNSFYFSEEMMSQIPHPKSIVVSDRATLFLNLMIG 240
 NPLA + L+M DL D+PYLSYDQG+HNSFYFSEEMMSQ+PH KSIVVSDRATLFLNLMIG
 20 Sbjet: 181 NPLATKSLLSMDLDRDFPYLSYDQGIHNSFYFSEEMMSQMPHNKSIVVSDRATLFLNLMIG 240

Query: 241 LDGYTVATGILNSKLNDEIVAIPLDVDDVIDIVYIRHDKANLSKMGQKFIDYLLLEEVSPN 301
 LDGYTVA+GILNS LNGD+IVAIPLDV D IDIV+I+H+KANLSKMG++FI+YLLLEV+F+
 25 Sbjet: 241 LDGYTVASGILNSNLNGDQIVAIPLDVDPDEIDIVFIKHEKANLSKMGERFIEYLLLEVTFD 301

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

25 Example 905

A DNA sequence (GBSx0960) was identified in *S.galactiae* <SEQ ID 2761> which encodes the amino acid sequence <SEQ ID 2762>. This protein is predicted to be CpsX protein. Analysis of this protein sequence reveals the following:

Possible site: 32
 30 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -14.91 Transmembrane 22 - 38 (13 - 42)
 INTEGRAL Likelihood = -14.65 Transmembrane 52 - 68 (44 - 77)
 INTEGRAL Likelihood = -6.74 Transmembrane 76 - 92 (73 - 97)

35 ----- Final Results -----
 bacterial membrane --- Certainty=0.6965(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

40 The protein has homology with the following sequences in the databases:

>GP:AAC44935 GB:U56901 putative transcriptional regulator [Bacillus subtilis]
 Identities = 120/389 (30%), Positives = 196/389 (49%), Gaps = 17/389 (4%)

Query: 2 KIGKKIVLMFTAIVLTTLVLAGVYLTSAITFSTGELSKTFKDFSTSSNKSDAIK-QTRAF 60
 KI K+I+L+F A+ L V+ LG Y + E + S+ +++ + + + F
 45 Sbjet: 19 KILKRIMLLF-ALALLVVVGLGGYKLYKTINADESVDALSRGNKSNLRNEVDMKKKPF 77

Query: 61 SILLMGVDTGSSERASKWEGNSDSMILVTVPKTKTMTSLERDTLTLTSGPKNNEMNG 120
 SIL MG++ +++ +G SDS+I+VT++PK K M S+ RDT L+G + G
 50 Sbjet: 78 SILFMGIEDYATKGQ---KGRSDSLIVVTLDPKNKTMKLSIPRDRVQLAG---DITG 130

Query: 121 VEAKLNAAYAAGGAQMAIMTVQDLLNITIDNYVQINMQGLIDLNVAVGGITVTNEFDFFPI 180
 + K+NAAY+ GG + TV++ L I ID YV ++ G D++N VGGI V FDF
 55 Sbjet: 131 SKTKINAAYSKGKDETVETVENFLQIPIDKYVTVDGFGKDVINEVGGIDVDVFFDFDE 190

Query: 181 SIAENEPEYQATVAPGTHKINGEQALVYARMRYDDPEGDYGRQKRQREVIQKVLKKILAL 240
 +E + + G +NGE+AL YARMR D GD+GR RQ++++ ++ ++ +
 60 Sbjet: 191 KSDVDESK-RIYFKKGEMHLNGEALAYARMKQDKRGDFGRNDRQKQILNALIDRMSSA 249

Query: 241 DSISSYRKILSAVSSNMQTINIEISSRTIPSLGYRDALRTIKTYQLKGEDATLSGGGSYQ 300
 +I+ KI S N++TNI I+ + + I T + G D L +Y

-1004-

Sbjct: 250 SNIAKIDKIAEKASENVETNIRITEGLALQOIYSGFTSKIDTILSITGSDLYLGPNNITYY 309

Query: 301 IVTSNHLLLEIQNRIRTELGLHKVNQLKTNATVYENLYGSTKSQTVNNNYDSSGQAPSYSD 360

LE ++R L H ++ +T T S + + + S+G +

5 Sbjct: 310 FEPDATNLE---KVRKTLQEH-LDYTPDTSTGTSGTEDGTDSSSSSGSTGTGTTTDTGTT 365

Query: 361 SHSSYANYSSGVDITGQSASTDQDSTASSH 389

+ SSY+N SS T + ST +T SS+

10 Sbjct: 366 NGSSYSNDSS---TSSNNSTNTNSTDSSY 391

There is also homology to SEQ ID 2764.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 906

15 A DNA sequence (GBSx0961) was identified in *S.agalactiae* <SEQ ID 2765> which encodes the amino acid sequence <SEQ ID 2766>. This protein is predicted to be CpsIaB. Analysis of this protein sequence reveals the following:

Possible site: 41

>>> Seems to have no N-terminal signal sequence

20 INTEGRAL Likelihood = -0.75 Transmembrane 121 - 137 (121 - 137)

----- Final Results -----

bacterial membrane --- Certainty=0.1298(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

25 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9877> which encodes amino acid sequence <SEQ ID 9878> was also identified.

No corresponding DNA sequence was identified in *S.pyogenes*.

30 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 907

A DNA sequence (GBSx0962) was identified in *S.agalactiae* <SEQ ID 2767> which encodes the amino acid sequence <SEQ ID 2768>. This protein is predicted to be cpsb protein. Analysis of this protein sequence reveals the following:

Possible site: 35

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -9.02 Transmembrane 182 - 198 (179 - 204)

40 INTEGRAL Likelihood = -5.57 Transmembrane 30 - 46 (24 - 48)

----- Final Results -----

bacterial membrane --- Certainty=0.4609(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

45 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 10785> and protein <SEQ ID 10786> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 9

-1005-

McG: Discrim Score: -8.96
 GvH: Signal Score (-7.5): 0.11
 Possible site: 35
 >>> Seems to have no N-terminal signal sequence
 5 ALOM program count: 2 value: -9.02 threshold: 0.0
 INTEGRAL Likelihood = -9.02 Transmembrane 182 - 198 (179 - 204)
 INTEGRAL Likelihood = -5.57 Transmembrane 30 - 46 (24 - 48)
 PERIPHERAL Likelihood = 6.21 113
 modified ALOM score: 2.30
 10
 *** Reasoning Step: 3
 ----- Final Results -----
 15 bacterial membrane --- Certainty=0.4609(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

20 Example 908

A DNA sequence (GBSx0963) was identified in *S.agalactiae* <SEQ ID 2769> which encodes the amino acid sequence <SEQ ID 2770>. This protein is predicted to be CpsIaD. Analysis of this protein sequence reveals the following:

Possible site: 61
 25 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -2.44 Transmembrane 149 - 165 (149 - 166)
 ----- Final Results -----
 30 bacterial membrane --- Certainty=0.1977(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 909

A DNA sequence (GBSx0964) was identified in *S.agalactiae* <SEQ ID 2771> which encodes the amino acid sequence <SEQ ID 2772>. Analysis of this protein sequence reveals the following:

Possible site: 25
 40 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -12.26 Transmembrane 276 - 292 (270 - 297)
 INTEGRAL Likelihood = -4.62 Transmembrane 10 - 26 (9 - 28)
 INTEGRAL Likelihood = -4.14 Transmembrane 41 - 57 (39 - 58)
 INTEGRAL Likelihood = -3.24 Transmembrane 100 - 116 (100 - 116)
 45 INTEGRAL Likelihood = -3.08 Transmembrane 445 - 461 (443 - 461)
 ----- Final Results -----
 50 bacterial membrane --- Certainty=0.5904(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8687> and protein <SEQ ID 8688> were also identified. Analysis of this protein sequence reveals the following:

-1006-

Lipop: Possible site: -1 Crend: 8
 McG: Discrim Score: 5.69
 GvH: Signal Score (-7.5): -5.63
 Possible site: 25
 5 >>> Seems to have an uncleavable N-term signal seq
 ALOM program count: 5 value: -12.26 threshold: 0.0
 INTEGRAL Likelihood = -12.26 Transmembrane 276 - 292 (270 - 297)
 INTEGRAL Likelihood = -4.62 Transmembrane 10 - 26 (9 - 28)
 INTEGRAL Likelihood = -4.14 Transmembrane 41 - 57 (39 - 58)
 10 INTEGRAL Likelihood = -3.24 Transmembrane 100 - 116 (100 - 116)
 INTEGRAL Likelihood = -3.08 Transmembrane 445 - 461 (443 - 461)
 PERIPHERAL Likelihood = 2.23 221
 modified ALOM score: 2.95
 15 *** Reasoning Step: 3
 ----- Final Results -----
 bacterial membrane --- Certainty=0.5904(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 20 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 910

25 A DNA sequence (GBSx0965) was identified in *S.agalactiae* <SEQ ID 2773> which encodes the amino acid sequence <SEQ ID 2774>. This protein is predicted to be CpsF. Analysis of this protein sequence reveals the following:

 Possible site: 13
 30 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -2.60 Transmembrane 79 - 95 (78 - 95)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.2041(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 35 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

40 Example 911

A DNA sequence (GBSx0966) was identified in *S.agalactiae* <SEQ ID 2775> which encodes the amino acid sequence <SEQ ID 2776>. This protein is predicted to be galactosyltransferase. Analysis of this protein sequence reveals the following:

45 Possible site: 39
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4634(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 50 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.pyogenes*.

-1007-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 912

A DNA sequence (GBSx0967) was identified in *S. galactiae* <SEQ ID 2777> which encodes the amino acid sequence <SEQ ID 2778>. Analysis of this protein sequence reveals the following:

```

Possible site: 23
>>> Seems to have an uncleavable N-term signal seq
  INTEGRAL    Likelihood = -12.47    Transmembrane    59 - 75 ( 54 - 82)
  INTEGRAL    Likelihood = -10.88    Transmembrane   309 - 325 ( 307 - 332)
10  INTEGRAL    Likelihood = -8.07    Transmembrane    33 - 49 ( 28 - 53)
  INTEGRAL    Likelihood = -6.48    Transmembrane   195 - 211 ( 187 - 212)
  INTEGRAL    Likelihood = -6.16    Transmembrane   285 - 301 ( 283 - 306)
  INTEGRAL    Likelihood = -4.09    Transmembrane   222 - 238 ( 221 - 240)
  INTEGRAL    Likelihood = -3.50    Transmembrane    78 - 94 ( 77 - 96)
15  INTEGRAL    Likelihood = -2.71    Transmembrane   101 - 117 ( 99 - 117)
  INTEGRAL    Likelihood = -2.44    Transmembrane     8 - 24 ( 7 - 25)
  INTEGRAL    Likelihood = -1.59    Transmembrane   147 - 163 ( 147 - 164)
  INTEGRAL    Likelihood = -0.48    Transmembrane   168 - 184 ( 168 - 184)

20  ----- Final Results -----
      bacterial membrane --- Certainty=0.5989(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

25  The protein has homology with the following sequences in the GENPEPT database.

  >GP:CAB43614 GB:AJ239004 polysaccharide polymerase [Streptococcus pneumoniae]
    Identities = 74/309 (23%), Positives = 137/309 (43%), Gaps = 36/309 (11%)

30  Query: 53  FERRKLIV---IIFLLFIATILNLFVHKVTFILTLIFFLALKDI--SLKKAFSIIIGSRI 107
      FE+RK    II ++ I T+L    +    ++    +F+ + I    L++    II
  Sbjct: 61  FEKRKYTLQFIISIILITLTLTYTSIQMQNYVYFTSWFMLIGTIHYDLRRVIKIIIFIVS- 119

  Query: 108  LGVLLNQIFVKLDLIEIKY-----VNFYRDGQFILRSDLGFGHPNFIHNFALTIFLYIV 162
      L ++    IF+ L + I Y    +N R+ + +    GF HPN    +    ++I
35  Sbjct: 120  LSIMFISIFISLIMYIIDYKREILINIRRN-ETVRAFTFGFIHPNKFTIVLSNLCIMFIW 178

  Query: 163  LNYKRLKPVVMVFLTLNLYLLYQYTFSRITGYIIVILFIVLIYVTKNSLIKRVMFKLAPYV 222
      L    RLK    +    L +    Y +T +RT    + I+    L+Y+    ++ + ++    Y
40  Sbjct: 179  LIKDRLKYYHVTFCLETLQFFYFFFTQTRTALLVSIIVIFALLYI--YMFVENLELRWIGYS 236

  Query: 223  QFFLLVFTFLSSTIFFNSN--FVQKLDVLLTGRLHY-AHLQLVDGLTPFGNSFKE----- 274
      F + F + + F+ SN F    +D +LTGR+    A+ +    G T +G    +
45  Sbjct: 237  FFCISTFLGVLAQFYPSNNKFSIFIDNILTGRIKLAAYARTFFGYTFWGYVDKEIVWD 296

  Query: 275  -----TSVLFDNSYSMLLSMYGVVLTMFMIY-----YTIYSKKIIIIELQLLLFIMSII 324
      TS    FD+ YS L+S    G++    +    +++    Y+ +K +I+    LL + M    +
50  Sbjct: 297  PIWGLTSFTFDSFYFILMSNAGIIWLLILSVLFVKLQKYLDNKSLLI-----LLAWSMYAV 352

  Query: 325  LFTESFYPS 333
      T+    +PS
  Sbjct: 353  TETDLIFPS 361

```

No corresponding DNA sequence was identified in *S. pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-1008-

Example 913

A DNA sequence (GBSx0968) was identified in *S.agalactiae* <SEQ ID 2779> which encodes the amino acid sequence <SEQ ID 2780>. This protein is predicted to be cap8J. Analysis of this protein sequence reveals the following:

```

5   Possible site: 57
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
10      bacterial cytoplasm --- Certainty=0.3424(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15   >GP:CAB43613 GB:AJ239004 cap8J [Streptococcus pneumoniae]
      Identities = 94/237 (39%), Positives = 135/237 (56%), Gaps = 10/237 (4%)

      Query: 1   MIPKVIHYCWFGGNELPDNLKKYIKTWREQCPDYEIIEWNEHNYDVSKNVFMREAYTKKN 60
                  MIPK IHY WFGG+ PD + K I +W++ PDYEI+EWNE N+D+S + F + AY +
20   Sbjct: 1   MIPKKIHYIWFGGSEKPDVVLKICINSWKKYMPDYEIVEWNEFNFDLSDSQFAKSAYESRK 60

      Query: 61   FAYVSDYARLDIIYTYGGFYLDTDVELLKSL-DPLRIHECFIAREISCDVNTGLIIGAVK 119
                  +A+ SDYAR I+ YGG Y DTDVELLK++ D + H F E +VN GL+ +
20   Sbjct: 61   WAFASDYARFKILSKYGGIYFDTDVELLKTISDDILAHSSFTGFEYIGEVPGLVYACMP 120

      Query: 120  GHHFLKSNMSIYDKS--DLTSLNKTCEVITNLLINRGLKNKNIQKIDDITIYPRNYFN 177
                  K + Y+++ D+ L T + T+ L+ + N Q ID + IYP +YF
25   Sbjct: 121  DDKIAKYMVQYEQASFDINHL-VTVNTIITDYLLKNNFQKNNQFQIIDGLAIYDDYFC 179

      Query: 178  PKNLLTGKVDCLTSVTYSIHHYEGSWKSSSFISDSLKIRVRLIIDFLFGYGYRMLL 234
                  + +V LT T SIHHY +WK+ +LK +V++I+ + G YR LL
30   Sbjct: 180  GYDQEVKEVR-LTERTISIHHSATWKTR-----TLKRKVQMIVKTIIGAENYRKLL 230

```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 914

A DNA sequence (GBSx0969) was identified in *S.agalactiae* <SEQ ID 2781> which encodes the amino acid sequence <SEQ ID 2782>. Analysis of this protein sequence reveals the following:

```

40   Possible site: 44
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
45      bacterial cytoplasm --- Certainty=0.3897(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

50   >GP:CAA87700 GB:Z47767 WbcL [Yersinia enterocolitica]
      Identities = 60/207 (28%), Positives = 101/207 (47%), Gaps = 22/207 (10%)

      Query: 4   IFTPTFNRGYRLSYLYDSLNCQNTKNFIWLIVDDGSEDSTKEIVSNYIKENKVSIVYLYK 63
                  +FTPTFNR + L Y S+ Q + WLIVDDGS D+T E+V ++ ENK++I Y+Y+
50   Sbjct: 6   VFTPTFNRAHVLKRCYLSILEQDRDDIEWLIVDDGSTDNIAEVDVDSFKIENKLNKIYIYQ 65

      Query: 64   RNGGKHSAYNLAMRYMQPSDYHVCVDSDDWLLLEDAV-----BIIFKDLESITLSNRYVG 117
                  N GK +A+N A+ +Y + +DSDD + ++ +F D E + +
55

```

-1009-

Sbjct: 66 DNSGKQAAWNKAVENAS-GEYFIGLSDDDAFTAGSINKLLSMNAVFDDEKEIIGIR----A 120

Query: 118 LVYPRYSLNQGNWLNPKILEVNI PDLKYKYLKIETCIVINNAYLVDFEFPCEGENFL 177
 + +L N +L+ + + + D ++ ++ E L + +P G NF+

5 Sbjct: 121 ISVSSETLKPNNYYLSNEDKKSSWFD-EFSSGIRGERIDFFKTELLRKLYLPVASGINFI 179

Query: 178 SEEIMYIYLSKKGYFCPQNRKIYCFDY 204
 E Y ++K+ YCF Y

10 Sbjct: 180 PEIWFYSTVAKE-----YCFYY 196

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 915

15 A DNA sequence (GBSx0970) was identified in *S.galactiae* <SEQ ID 2783> which encodes the amino acid sequence <SEQ ID 2784>. This protein is predicted to be eps7. Analysis of this protein sequence reveals the following:

Possible site: 32
 >>> Seems to have an uncleavable N-term signal seq

20 INTEGRAL Likelihood = -2.18 Transmembrane 190 - 206 (189 - 206)

----- Final Results -----
 bacterial membrane --- Certainty=0.1871(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 25 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB59293 GB:AJ131984 putative galactosyl transferase
 [Streptococcus pneumoniae]

30 Identities = 101/312 (32%), Positives = 172/312 (54%), Gaps = 4/312 (1%)

Query: 3 LISIIVPVYNGEIVIGRCLDSILEQTYQNLEIIIDDGSSDRITGDICEKYFLEDRIKIF 62
 +IS+IVPVYN Y+ LDS+LEQTY++ E+I+++DGS+D +G+IC++Y I F

35 Sbjct: 1 MISVIVPVYNVADYLRFAIDSLLEQTYKDFEVILVNDGSTDNGEICDEYKLYDNIHVF 60

Query: 63 YQENRGQSVARNNGVLRCTGDWIAFLDSDVYLPYSIEVMYNIQKATNADIVLT--SIGN 120
 +++N G S ARN G+ + G++I FLSDDD + PY++E++ IQK + DIV T I

Sbjct: 61 HKKNGGLSDARNFGLEKSRGEFITFLDSDDYFEPYALELLITIQKDYDVIDIVSTKGGITY 120

40 Query: 121 FNNTYNTSINSQYLKEIKLYTLEVALEEMYYGKTYGVSPLAKLYPRSNLLSNPYPEGKIH 180
 ++ Y+ + ++ +K+ T + L +YY VS KLY R +L +P+GKI+

Sbjct: 121 SHDIYSKKLMAEDYLTIVKILTNEFLAAVYNDEMTVSAWGKLYKR-DLFKTIFFPKGIY 179

45 Query: 181 EDMDTFFKLISCASKIAVCDIVTAVVYFSDNSTTRTKFNERMLYFFEAIQNNIVFINLNF 240
 ED+ + + +A D+ Y S + F++R FF+AI +N I +

Sbjct: 180 EDLYVVAERLLNIKTVAHTDLPYHYRQGSIVNSTFSDRQYDFDAIDHNEAIKKFY 239

Query: 241 PHNTSLISAVIYNEVFGGIDICGKMIDFKLYDITVDYRKKYRKYFKTILFNNRISVKEKV 300
 + L++A+ V G I + + + + Y+ ++ N +I +K KV

50 Sbjct: 240 CGDKELLAALNAKRIVIGSF-ILSNSAFYNSKNDITKIIRIIPYWEVIKKNKIPMKRKV 298

Query: 301 KYILFISSIRYF 312
 + +LF+ S Y+

55 Sbjct: 299 QCVLFLLSPNY 310

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-1010-

Example 916

A DNA sequence (GBSx0971) was identified in *S.agalactiae* <SEQ ID 2785> which encodes the amino acid sequence <SEQ ID 2786>. This protein is predicted to be galactosyltransferase. Analysis of this protein sequence reveals the following:

```

5   Possible site: 26
    >>> Seems to have an uncleavable N-term signal seq

    ----- Final Results -----
10          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2787> which encodes the amino acid sequence <SEQ ID 2788>. Analysis of this protein sequence reveals the following:

```

15   Possible site: 28
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
20          bacterial cytoplasm --- Certainty=0.2065(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

25   Identities = 37/111 (33%), Positives = 61/111 (54%), Gaps = 3/111 (2%)

Query: 1   MDKVSIIIPVYNVQSFLNECIESVLAQ-TYSNLEIILVNDGSTDNSGDIC-DYYSEIDGR 58
          M KVSII  YN  ++++ ++S L+Q T  +EII+++D STD+S +I  Y  +  G+
Sbjct: 1   MYKVSIIICTNYNKAPWISDALDSFLSQVTDFFEVEIIVIDDASTDDSREILKSYQKKSSGK 60

30   Query: 59  I-FVFHKNNGGLSDARNYGISRATGDYIYLLDSDDYLYKEDAIERMVEFSE 108
          I  +F++ N G++          A G YI  D DDY  +++ V+  E
          Sbjct: 61  IKLLFNETNIGITKTKTWIKACLYAKGKYIARCDGDDYWTDSFKLQKQVDVLE 111

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 917

A DNA sequence (GBSx0972) was identified in *S.agalactiae* <SEQ ID 2789> which encodes the amino acid sequence <SEQ ID 2790>. This protein is predicted to be CpsK. Analysis of this protein sequence reveals the following:

```

40   Possible site: 52
    >>> Seems to have an uncleavable N-term signal seq

    ----- Final Results -----
45          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-1011-

Example 918

A DNA sequence (GBSx0973) was identified in *S.agalactiae* <SEQ ID 2791> which encodes the amino acid sequence <SEQ ID 2792>. Analysis of this protein sequence reveals the following:

```

5      Possible site: 31
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.1956(Affirmative) < succ>
10     bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 919

A DNA sequence (GBSx0974) was identified in *S.agalactiae* <SEQ ID 2793> which encodes the amino acid sequence <SEQ ID 2794>. This protein is predicted to be capsular polysaccharide. Analysis of this protein sequence reveals the following:

```

20     Possible site: 36
      >>> Seems to have an uncleavable N-term signal seq

      INTEGRAL    Likelihood = -8.81    Transmembrane    89 - 105 ( 80 - 112)
      INTEGRAL    Likelihood = -7.01    Transmembrane   439 - 455 ( 428 - 460)
      INTEGRAL    Likelihood = -6.74    Transmembrane   322 - 338 ( 317 - 342)
25     INTEGRAL    Likelihood = -4.88    Transmembrane   175 - 191 ( 174 - 195)
      INTEGRAL    Likelihood = -3.45    Transmembrane   146 - 162 ( 145 - 166)
      INTEGRAL    Likelihood = -3.08    Transmembrane   381 - 397 ( 375 - 398)
      INTEGRAL    Likelihood = -2.50    Transmembrane   413 - 429 ( 412 - 430)
      INTEGRAL    Likelihood = -1.91    Transmembrane   206 - 222 ( 205 - 222)
30     INTEGRAL    Likelihood = -1.59    Transmembrane   354 - 370 ( 354 - 372)
      INTEGRAL    Likelihood = -1.54    Transmembrane    43 - 59 ( 43 - 61)
      INTEGRAL    Likelihood = -0.22    Transmembrane   252 - 268 ( 252 - 268)

      ----- Final Results -----
35     bacterial membrane --- Certainty=0.4524(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 920

A DNA sequence (GBSx0975) was identified in *S.agalactiae* <SEQ ID 2795> which encodes the amino acid sequence <SEQ ID 2796>. This protein is predicted to be NeuB. Analysis of this protein sequence reveals the following:

```

50     Possible site: 30
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2992(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

```

-1012-

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 921

A DNA sequence (GBSx0976) was identified in *S.agalactiae* <SEQ ID 2797> which encodes the amino acid sequence <SEQ ID 2798>. This protein is predicted to be NeuC. Analysis of this protein sequence reveals the following:

10 Possible site: 41
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
 15 bacterial cytoplasm --- Certainty=0.3150(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 922

A DNA sequence (GBSx0977) was identified in *S.agalactiae* <SEQ ID 2799> which encodes the amino acid sequence <SEQ ID 2800>. This protein is predicted to be neuD. Analysis of this protein sequence reveals the following:

25 Possible site: 16
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
 30 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

There is homology to SEQ ID 542.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 923

A DNA sequence (GBSx0979) was identified in *S.agalactiae* <SEQ ID 2801> which encodes the amino acid sequence <SEQ ID 2802>. Analysis of this protein sequence reveals the following:

40 Possible site: 33
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
 45 bacterial cytoplasm --- Certainty=0.2576(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 924

A DNA sequence (GBSx0980) was identified in *S.agalactiae* <SEQ ID 2803> which encodes the amino acid sequence <SEQ ID 2804>. Analysis of this protein sequence reveals the following:

Possible site: 49
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1621(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9875> which encodes amino acid sequence <SEQ ID 9876> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2805> which encodes the amino acid sequence <SEQ ID 2806>. Analysis of this protein sequence reveals the following:

Possible site: 51
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1066(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 83/139 (59%), Positives = 111/139 (79%)

Query: 6 TETHDHOALIQKLLVSIHYLTLFRDEIILVEKTPSLGKHFSIAIVQNELGEILSKIEAL 65
TE + HQ LIQKLLVSIHYLTLFRDE+ LVE+TPS+LG F +VQ+ELG+I++ I+ L
Sbjct: 4 TEQNSHQILIQKLLVSIHYLTLFRDELKLVERTPSILGGEFPAHLVQSELGDIVAAIDTL 63

Query: 66 SKQKQLIRSIYWYDESSFKVMNKALAIIVEEWIKGLDNLLEFCQSQTVFQAILGDERAHVF 125
Q++LI S +WY+ES+FK+MNK L IV+ WIKG+D+L++ CQS+ VFQ I+GD+R VF
Sbjct: 64 DMQQRLESTFWYEESAFKLMNKTLDIVDNWIKGVHDHLDLCQSKEVFIIGDKRIRVF 123

Query: 126 GILIDVYTSINIINTSLKE 144
G+L DV++SL + SLKE
Sbjct: 124 GVLSDVFSLSLKVSAISLKE 142

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 925

A DNA sequence (GBSx0981) was identified in *S.agalactiae* <SEQ ID 2807> which encodes the amino acid sequence <SEQ ID 2808>. This protein is predicted to be uracil-DNA glycosylase (ung). Analysis of this protein sequence reveals the following:

Possible site: 34
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3427(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-1014-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2809> which encodes the amino acid sequence <SEQ ID 2810>. Analysis of this protein sequence reveals the following:

Possible site: 33
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4200(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 160/216 (74%), Positives = 185/216 (85%)

Query: 1 MKHSSWHDLIKRELPHNYNKNITFMDAVYESGIVYPPRDKVFNAIQITPLENVKVVIIIG 60
 M HS WH+ IK LP HY+ +IN F+D Y SG+VYPPR+ VF A+Q+TPLE KV+I+G
 Sbjct: 1 MAHSIWHEKIKSFLPEHYGRINHLDEAYASGLVYPPRENVFKALQVTPLEETKVLILG 60

Query: 61 QDPYHGPQQAQGLSFSVFDNLFAPPSLQNLKELAEIDIGSRSHDLTWSAQGVLLLNAC 120
 QDPYHGP+QAQGLSFSVP+ + APPSL NILKELA+DIG R HHDL++WA QGVLLLNAC
 Sbjct: 61 QDPYHGPQQAQGLSFSVPEISAPPSLINILKELADDIGPRDHDLTWSAQGVLLLNAC 120

Query: 121 LTVPEHQANGHAGLIWEFFTDAAVIKVVNQKETPVVFIWGGYARKKSLIDNPIHHIES 180
 LTVP QANGHAGLIWEFFTDAAVIKVN+K++PVVFIWLG YARKKK+ I NP HHIES
 Sbjct: 121 LTVPAGQANGHAGLIWEFFTDAAVIKVLNEKDSPVVFIWLGAYARKKAFITNPKHHIES 180

Query: 181 PHPSPLSAYRGFFGSRPFSRTNHFLEEENGINEIDWL 216
 PHPSPLS+YRGFFGS+PFSRTN LE+EG+ +DWL
 Sbjct: 181 PHPSPLSSYRGFFGSKPFSRTNAILEKEGTMTCVDWL 216

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 926

A DNA sequence (GBSx0982) was identified in *S.galactiae* <SEQ ID 2811> which encodes the amino acid sequence <SEQ ID 2812>. Analysis of this protein sequence reveals the following:

Possible site: 20
 >>> Seems to have an uncleavable N-term signal seq

INTEGRAL Likelihood = -11.15 Transmembrane 147 - 163 (109 - 166)
 INTEGRAL Likelihood = -8.92 Transmembrane 124 - 140 (109 - 146)
 INTEGRAL Likelihood = -6.16 Transmembrane 167 - 183 (166 - 186)
 INTEGRAL Likelihood = -4.67 Transmembrane 3 - 19 (1 - 23)
 INTEGRAL Likelihood = -3.98 Transmembrane 72 - 88 (64 - 92)
 INTEGRAL Likelihood = -1.06 Transmembrane 106 - 122 (105 - 122)
 INTEGRAL Likelihood = -0.90 Transmembrane 54 - 70 (54 - 70)

----- Final Results -----
 bacterial membrane --- Certainty=0.5458(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9873> which encodes amino acid sequence <SEQ ID 9874> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA91549 GB:Z67739 unidentified [Streptococcus pneumoniae]
 Identities = 134/212 (63%), Positives = 168/212 (79%)

Query: 1 MNIIIMIIIIAYLLGSIQTGLWIGKYFYQVNLRQFGSGNTGTTNTFRILGVKAGIVTLTID 60

-1015-

M I+++I+AYLLGSI +GLWIG+ F+Q+NLR+HGSGNTGTTNTFRILG KAG+ T ID
 Sbjct: 1 MITIVLLILAYLLGSIPSGLWIGQVFFQINLREHGSGNTGTTNTFRILGKKAGMATFVID 60
 Query: 61 ILKGTLATLPIIILGITTVPFFIGFFAIIGHTFPPIFAQFKGGKAVATSAGVLLGFAPSF 120
 KGTLATL+PII + VSP G A+IGHTFPPIFA FKGGKAVATSAGV+ GFAP F
 Sbjct: 61 FFKGTLATLLPIIFHLQGVSPILFGLLAVIGHTFPPIFAGFKGGKAVATSAGVIFGFAPIF 120
 Query: 121 FLYLLVIFLLTLYLFSMISLSSITVAVVGILSVLIFPLVGFILTDYDWIFTTVVILMALT 180
 LYL +IF LYL SMISLSS+T ++ ++ VL+FPL GFIL++YD++F +++ +A
 Sbjct: 121 CLYLAIFFGALYLGSMISLSSVTASIAAVIGVLLFPLFGFILSNYDFLFIATILALASL 180
 Query: 181 IIRHODNIKRIRKQENLVPFGLNLSKQKNK 212
 IIRH+DNI RI+ + ENLVP+GLNL+ Q K
 Sbjct: 181 IIRHKDNIARIKNTENLVPWGLNLTHQDPK 212

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2813> which encodes the amino acid sequence <SEQ ID 2814>. Analysis of this protein sequence reveals the following:

Possible site: 17
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -10.83 Transmembrane 194 - 210 (191 - 216)
 INTEGRAL Likelihood = -9.77 Transmembrane 146 - 162 (132 - 191)
 INTEGRAL Likelihood = -7.70 Transmembrane 165 - 181 (163 - 191)
 INTEGRAL Likelihood = -5.89 Transmembrane 23 - 39 (19 - 47)
 INTEGRAL Likelihood = -4.83 Transmembrane 95 - 111 (91 - 118)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.5331(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:CAA91549 GB:Z67739 unidentified [Streptococcus pneumoniae]
 Identities = 138/213 (64%), Positives = 166/213 (77%)
 Query: 28 MKLLLFITIAAYLLGSIPTGLWIGQYFYHINLREHGSGNTGTTNTFRILGVKAGTATLAI 87
 M ++ + +AYLLGSIP+GLWIGQ F+ INLREHGSGNTGTTNTFRILG KAG AT ID
 Sbjct: 1 MITIVLLILAYLLGSIPSGLWIGQVFFQINLREHGSGNTGTTNTFRILGKKAGMATFVID 60
 Query: 88 MFKGTLSILLPIIFGMTSISIAIGFFAVLGHFTFPIFANFKGGKAVATSAGVLLGFAPLY 147
 FKGTL+ LLPIIF + +S + G AV+GHTFPPIFA FKGGKAVATSAGV+ GFAP++
 Sbjct: 61 FFKGTLATLLPIIFHLQGVSPILFGLLAVIGHTFPPIFAGFKGGKAVATSAGVIFGFAPIF 120
 Query: 148 LFFLASIFVLVLYLFSMISLASVSAIVGVLSVLTFPAIHFLPNYDYFLTFIVILLAFI 207
 +LA IF LYL SMISL+SV ++I V+ VL FP F+L NYD+ I++ LA +
 Sbjct: 121 CLYLAIFFGALYLGSMISLSSVTASIAAVIGVLLFPLFGFILSNYDFLFIATILALASL 180
 Query: 208 IIRHKDNISRIKHHTENLIPWGLNLSKQVPEK 240
 IIRHKDNI+RIK+ TENL+PWGLNL+ Q PKK
 Sbjct: 181 IIRHKDNIARIKNTENLVPWGLNLTHQDPK 213

An alignment of the GAS and GBS proteins is shown below.

Identities = 143/212 (67%), Positives = 174/212 (81%)
 Query: 1 MNIIIMIIAYLLGSIQTGLWIGKYFYQVNLQHGSGNTGTTNTFRILGVKAGIVTLTID 60
 M +++ I IAYLLGSI TGLWIG+YFY +NLR+HGSGNTGTTNTFRILGVKAG TL ID
 Sbjct: 28 MKLLLFITIAAYLLGSIPTGLWIGQYFYHINLREHGSGNTGTTNTFRILGVKAGTATLAI 87
 Query: 61 ILKGTLATLPIIILGITTVPFFIGFFAIIGHTFPPIFAQFKGGKAVATSAGVLLGFAPSF 120
 + KGTL+ L+PII G+T++S IGFFA++GHTFPPIFA FKGGKAVATSAGVLLGFAP +
 Sbjct: 88 MFKGTLSILLPIIFGMTSISIAIGFFAVLGHFTFPIFANFKGGKAVATSAGVLLGFAPLY 147
 Query: 121 FLYLLVIFLLTLYLFSMISLSSITVAVVGILSVLIFPLVGFILTDYDWIFTTVVILMALT 180
 +L IF+L LYLFSMISL+S+ A+VG+LSVL FP + F+L +YD+ T +VIL+A
 Sbjct: 148 LFFLASIFVLVLYLFSMISLASVSAIVGVLSVLTFPAIHFLPNYDYFLTFIVILLAFI 207

-1016-

Query: 181 IIIRHQDNIRKIRKRQENLVPFGLNLSKQKNK 212
 IIIRH+DNI RI+ ENL+P+GLNLSKQ K
 Sbjet: 208 IIIRHKDNISRIKHHTENLIPWGLNLSKQVPK 239

5

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 927

10 A DNA sequence (GBSx0983) was identified in *Sagalactiae* <SEQ ID 2815> which encodes the amino acid sequence <SEQ ID 2816>. Analysis of this protein sequence reveals the following:

Possible site: 18
 >>> Seems to have an uncleavable N-term signal seq

15

----- Final Results -----

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

20 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 928

25 A DNA sequence (GBSx0984) was identified in *Sagalactiae* <SEQ ID 2817> which encodes the amino acid sequence <SEQ ID 2818>. Analysis of this protein sequence reveals the following:

Possible site: 56
 >>> Seems to have no N-terminal signal sequence

30

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1585(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

35 A related GBS nucleic acid sequence <SEQ ID 9871> which encodes amino acid sequence <SEQ ID 9872> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA91550 GB:Z67739 DNA topoisomerase IV [Streptococcus pneumoniae] (ver 2)
 Identities = 574/649 (88%), Positives = 617/649 (94%), Gaps = 2/649 (0%)

40

Query: 5 LAKQDITVTNYGDDAIQVLEGLDAVRKRPGMYIGSTDGTLHHLVWEIVDNAVDEALSGF 64
 ++K++I + NY DDAIQVLEGLDAVRKRPGMYIGSTDG GLHHLVWEIVDNAVDEALSGF
 Sbjet: 1 MSKKEININNYDDAIQVLEGLDAVRKRPGMYIGSTDGAGLHHLVWEIVDNAVDEALSGF 60

45

Query: 65 GNRIDVIINKDGSITVTDHGRGMPTGMHAMGKPTVEVIFTVLHAGGKFGQGGYKTSGGLH 124
 G+RIDV INKDGs+TV DHGRGMPTGMHAMG PTVEVIFT+LHAGGKFGQGGYKTSGGLH
 Sbjet: 61 GDRIDVTINKDGSITVQDHGRGMPTGMHAMGIPTVEVIFTILHAGGKFGQGGYKTSGGLH 120

50

Query: 125 GVGSSVVNALSSWLEVEIIRDGAIYRQRFENGKGPVTTLLKIGTAPKSKSGTSVSFMPDQ 184
 GVGSSVVNALSSWLEVEI RDGA+Y+QRFENGKGPVTTLLKIGTAPKSK+GT V+FMPD
 Sbjet: 121 GVGSSVVNALSSWLEVEITRDGAVYKQRFENGKGPVTTLLKIGTAPKSKTGTKVTFMPDA 180

Query: 185 SVFSTIDFKFNTIAERLKESAFLLKNVTLLTLDNRSEEAHLEPHYENGVDVFEYLNED 244
 ++FST DFK+NTI+ERL ESFLLKNVTLL+LTD R++EA +EPHYENGVDV YLNED

-1017-

Sbjct: 181 TIFSTTDFKYNTISERLINESAFLKKNVTLSLTDKRTDEA--IEFHYENGVDQFVSYL NED 238

Query: 245 KETLTPIMFFEGEEQEFHIEVALQYNDGFSNLSFVNNVVRTKDGGTHETGLKSAITKSM 304
KE LTP+++FEGE+ F +EVALQYNDGFSNLSFVNNVVRTKDGGTHETGLKSAITK M

5 Sbjct: 239 KEILTPVLYFEGEDNGFQVEVALQYNDGFSNLSFVNNVVRTKDGGTHETGLKSAITKVM 298

Query: 305 NDYARKTGLLKEKDKNLEGS DYREGLSAILSILVPEEHLQFEGQTKDKLGSPLARPIVDG 364
NDYARKTGLLKEKDKNLEGS DYREGL+A+LSILVPEEHLQFEGQTKDKLGSPLARP+VDG

10 Sbjct: 299 NDYARKTGLLKEKDKNLEGS DYREGLAAVL SILVPEEHLQFEGQTKDKLGSPLARPVVDG 358

Query: 365 IVSEKLT VFLMENGDLASNLIRKAIKARDAREAAARKARDES RNKGKSKDKGLLSGKLT P 424
IV++KLT+FLMENG+LASNLIRKAIKARDAREAAARKARDES RNKGK+KKDKGLLSGKLT P

15 Sbjct: 359 IVADKLT VFLMENGELASNLIRKAIKARDAREAAARKARDES RNKGKKNKDKGLLSGKLT P 418

Query: 425 AQSKNAKNELYLVEGDSAGGSAKQGRDRKFQAILPLRGKVLNTAKAKMADI+KNEEINT 484
AQSKN KNELYLVEGDSAGGSAKQGRDRKFQAILPLRGKV+NTAKAKMADI+KNEEINT

20 Sbjct: 419 AQSKNPAKNELYLVEGDSAGGSAKQGRDRKFQAILPLRGKVINTAKAKMADILKNEEINT 478

Query: 485 MIHTIGAGVGPDFNLDDIN YDKIIIMTDADTDGAHIQTL LLLTFFRYMRPLVEEGHVYIA 544
MI+TIGAGVG DF++D NYDKIIIMTDADTDGAHIQTL LLLTFFRYMRPLVE GHVYIA

25 Sbjct: 479 MIYTIGAGVGADFSIEDANYDKIIIMTDADTDGAHIQTL LLLTFFRYMRPLVEAGHVYIA 538

Query: 545 LPPLYKMSKKGKKEIVEYAWTDIELEELRQKFGKSGLLQRYKGLGEMNADQLWETTMNP 604
LPPLYKMSKKGKKE V YAWTD ELEELR++FGKG+ LQRYKGLGEMNADQLWETTMNP

30 Sbjct: 539 LPPLYKMSKKGKKEEVAYAWTDIELEELRQKFGKGATLQRYKGLGEMNADQLWETTMNP 598

Query: 605 ETRTLIRVTIEDLARAERRVNLMDGKVP RRQWIEDNVKFTLEENTVF 653
ETRTLIRVTIEDLARAERRVNLMDGKV PRR+WIEDNVKFTLEE TVF

35 Sbjct: 599 ETRTLIRVTIEDLARAERRVNLMDGKVEPRRKWIEDNVKFTLEEATVF 647

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2819> which encodes the amino acid sequence <SEQ ID 2820>. Analysis of this protein sequence reveals the following:

Possible site: 53

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1518(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 560/649 (86%), Positives = 615/649 (94%)

Query: 5 LAKQDITVTNYGDDAIQVLEGLDAVRKRP GMYIGSTDGTGLHHLVWEIVDNAVDEALSGF 64
L K++IT+ NY DDAIQVLEGLDAVRKRP GMYIGSTD TGLHHL+WEIVDNAVDEALSGF

45 Sbjct: 2 LTKKEITINNYNDDAIQVLEGLDAVRKRP GMYIGSTDATGLHHLIWEIVDNAVDEALSGF 61

Query: 65 GNRIDVIINKDGSITVTDHGRGMP TGMHAMGKPTVEVIFTVLHAGGKFGQGGYKTS GGLH 124
G+ I V+INKDGS++V D GRGMP TGM HAMG PTV+VIFT+LHAGGKFGQGGYKTS GGLH

50 Sbjct: 62 GDDIKVVINKDGSVSVADSGRGMP TGMHAMGIPTVQVIFTILHAGGKFGQGGYKTS GGLH 121

Query: 125 GVGSSVVNALSSWLEVEIIRDGAIRQR FENGKGPVTTLKKIGTAPKSKSGTSV SFMPDQ 184
GVGSSVVNALSSWLEVEI RDG++YRQR FENGKGPVTTLKK+GTAPKSKSGT V+FM PD

55 Sbjct: 122 GVGSSVVNALSAWLEVEITRDGSVYRQR FENGKGPVTTLKKVGTAPKSKSGTVTFMPDD 181

Query: 185 SVFSTIDFKENTIAERLKESAFLLKKNVTLTLDNRSEEAHLEFHYENGVDQFVEYL NED 244
+FSTIDFKENT I+ERLKESAFLLKNV ++LTD R ++ E FHYENGVDQFVEYL NED

60 Sbjct: 182 KIFSTIDFKENTISERLKESAFLLKKNVMSLTDLRGDDPIIEEFHYENGVDQFVEYL NED 241

Query: 245 KETLTPIMFFEGEEQEFHIEVALQYNDGFSNLSFVNNVVRTKDGGTHETGLKSAITKSM 304
KETLTP+++ EG++Q+F +EVALQYNDGFSNLSFVNNVVRTKDGG+HETGLKSAITK+M

65 Sbjct: 242 KETLTPVIYMEGQDQDFQVEVALQYNDGFSNLSFVNNVVRTKDGGSHETGLKSAITKAM 301

Query: 305 NDYARKTGLLKEKDKNLEGS DYREGLSAILSILVPEEHLQFEGQTKDKLGSPLARPIVDG 364
NDYARKT LLKEKDKNLEGS DYREGLSA+LSILVPE+HLQFEGQTKDKLGSPLARPIV+

-1018-

Sbjct: 302 NDYARKTNLLKEKDKNLEGS DYREGLSAVLSILVPEQHLQFEGQTKDKLGSPLARPIVES 361

Query: 365 IVSEKLT YFLMENGDLASNLIRKAIKARDAREAAARKARDES RNGKSKKDKGLLSGKLT P 424
IVSEKLT+FL+ENG++AS+L+RKAIKARDAREAAARKARD+SRNGKK+KKDKGLLSGKLT P

5 Sbjct: 362 IVSEKLTFFLLENGEVASHLVRKAIKARDAREAAARKARDDSRNGKKNKKDKGLLSGKLT P 421

Query: 425 AQSKNAKKNELYLVEGDSAGGSAKQGRDRKFQAILPLRGKVLNTAKAKMADI I KNEEINT 484
AQSKNAKKNELYLVEGDSAGGSAKQGRDRKFQAILPLRGKVLNT KAKMADI+KNEEINT

10 Sbjct: 422 AQSKNAKKNELYLVEGDSAGGSAKQGRDRKFQAILPLRGKVLNTEKAKMADILKNEEINT 481

Query: 485 MIHTIGAGVGPDFNLDDIN YDKIIIMTDADTDGAHIQTLLLTFFRYMRPLVEEGHVYIA 544
M++TIGAGVG DFN L+DIN YDKIIIMTDADTDGAHIQTLLLTFFRYMRPLVE GHVYIA

15 Sbjct: 482 MVYTIGAGVGADFNLEDIN YDKIIIMTDADTDGAHIQTLLLTFFRYMRPLVEAGHVYIA 541

Query: 545 LPPLYKMSKGKKGKBEIVEYAWTDIELEELRQKFGKGSLLQRYKGLGEMNADQLWETTMNP 604
LPPLYKMSKGKKGK E + YAWTD ELE+LR++FGKG++LQRYKGLGEMNA+QLWETTM+P

20 Sbjct: 542 LPPLYKMSKGKKGKTEKIAYAWTDGELEDLRREFGKGAILQRYKGLGEMNANQLWETTMNDP 601

Query: 605 ETRTLIRVTIEDLARAERRVNVLMGDKVPPRRQWIEDNVKFTLEENTVF 653
ETRTLIRVTI+DLARAERRV+VLMGDK PRRQWIEDNVKFTLEENTVF

Sbjct: 602 ETRTLIRVTIDDLARAERRVSVLMGDKAAPRRQWIEDNVKFTLEENTVF 650

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

25 Example 929

A DNA sequence (GBSx0985) was identified in *Sagalactiae* <SEQ ID 2821> which encodes the amino acid sequence <SEQ ID 2822>. Analysis of this protein sequence reveals the following:

Possible site: 49
>>> Seems to have no N-terminal signal sequence

30 INTEGRAL Likelihood = -0.80 Transmembrane 378 - 394 (378 - 394)

----- Final Results -----

35 bacterial membrane --- Certainty=0.1319(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAD34369 GB:AF129764 ParC [Streptococcus mitis]
Identities = 640/820 (78%), Positives = 722/820 (88%), Gaps = 5/820 (0%)

40 Query: 1 MSNIQNMSLEDIMGERFGRYSKYIIQERALPDIRDGLKPVQRRILYSMNKDGNTEFKGFR 60
MSNIQNMSLEDIMGERFGRYSKYIIQ+RALEPDIRDGLKPVQRRILYSMNKDGNTEF+K +R
Sbjct: 1 MSNIQNMSLEDIMGERFGRYSKYIIQDRALEPDIRDGLKPVQRRILYSMNKDGNTEFKSYR 60

45 Query: 61 KSAKSVGNVMGNFHPHGDSSIIDAMVRMSQDWKNRETLIEMHGNNGSMGDGPAAAMRYTE 120
KSAKSVGN+MGNFHPHGDSSIIDAMVRMSQDWKNRE L+EMHGNNGSMGDGP AAMRYTE
Sbjct: 61 KSAKSVGNIMGNFHPHGDSSIIDAMVRMSQDWKNREILVEMHGNNGSMGDGPPAAMRYTE 120

50 Query: 121 ARLSEIAGYLLQDIDKNTVPFAWNFDDETEKEPTVLPAAFPNLLVNGATGISAGYATDIPP 180
ARLSEIAGYLLQDIDK TVPF+WNFDDETEKEPTVLPAAFPNLLVNG+TGISAGYATDIPP
Sbjct: 121 ARLSEIAGYLLQDIDKKTVPFWSNFDDETEKEPTVLPAAFPNLLVNGSTGISAGYATDIPP 180

55 Query: 181 HNLAEVIDAVVYMIDHPKAKLDKLMFEFLPGDFPTGAI IQGKDEIRKAYETGKGRVAVRS 240
HNLAEVIDA VYMIDHP AK+DKLMFEFLPGDFPTG IIQG+DEI+KAYETGKGRV VRS
Sbjct: 181 HNLAEVIDAAVYMIDHPTAKVDKLMFEFLPGDFPTGGIIQGRDEIKKAYETGKGRVVRS 240

60 Query: 241 RTAIEITLKGKQIIVTEIPYEVNKSVLVKRIDDVRVNNKVPGIAEVRDES DRDGLRIAI 300
+T IE LKGKQ+QI++TEIPYE+NK+ LVK+IDDVRVN+KV GIAEVRDES DRDGLRIAI
Sbjct: 241 KTRIEIKLKGKQEVITEIPYEINKANLVKKIDDVRVNSKVAGIAEVRDES DRDGLRIAI 300

Query: 301 ELKKEADETIVLNLYFKYTDLQVNYNFMVAIDDYTPKQVGLSRILTSYIAHRREIIAR 360
ELKK+A+ +VLNLYFKYTDLQ+NYNFMVAID++TP+QVG+ IL+SYIAHRRE+I+AR

-1019-

Sbjct: 301 ELKKDANTELVNLNFKYTDLQINYNFNMVAIDNFTFRQVGIVPILSSYIAHRREVILAR 360

Query: 361 SKFDKEKAEKRLHIVEGLIRVLSILDEVIALIRASENKADAKENLKVSVEFSEAQAIAIV 420
S+FDKEKAEKRLHIVEGLIRV+LILDEVIALIRASENKADAKENLKVS+Y+QAEIAIV

5 Sbjct: 361 SRFDKEKAEKRLHIVEGLIRVISILDEVIALIRASENKADAKENLKVSDFTEEQAEIAIV 420

Query: 421 TLQLYRLTNIDIVTLREEEELRQQITMLKAIISDERTMYNVMKRELREVKKKFANTRRS 480
TLQLYRLTNID+V L+EEE ELR++I ML AII DERTMYN+MK+ELREVKKKFA R S

10 Sbjct: 421 TLQLYRLTNIDVVVLQEEEAELREKIAMLAIIIGDERTMYNLMKKELREVKKKFATPRLS 480

Query: 481 ELQELAEITIEDTASLIIEEDTYVSVTRGGYVKRTSPRSFNASTVDELGKREDELIFVS 540
L++ A+ IEIDTASLI EEDTYVSVT+ GY+KRTSPRSF AST++E+GKR+DD LIFV

Sbjct: 481 SLEDTAKAIEIDTASLIAEEDTYVSVTKAGYIKRTSPRSFAASTLEEIGKRDDRLIFVQ 540

15 Query: 541 NAKTTQHLLMFTNLGNLAYRPVHELADIRWKDVGEHLSQNLVNFASNEEIIYAEVDDF- 599
+AKTTQHLLMFT LGN+ YRP+HELADIRWKD+GEHLSQ + NF +NEEI+Y E+VD F

Sbjct: 541 SAKTTQHLLMFTTLGNVIYRPIHELADIRWKDIGEHLSTITNFETNEEIIYVEVVDQFD 600

Query: 600 TKETYFAVTSLGQIKRFERQEISPWRTYKSKTAKYAKLKSVEDYVVTVAPIQLEDVILVT 659
TYFA T LGQIKR ER+E +PWRTYKSK+ KYAKLK D +V VAPI+L+DV+L++

20 Sbjct: 601 DATTTYFAATRLGQIKRVERKEFTPWRTYKSKSVKYAKLKDDTDQIVAVAPIKLDDVLLIS 660

Query: 660 YNGYALRFSINDVPVVGSKAAGVKAMNLDKRDHIVSAFIANTTSYLLTHRGSLKRMAD 719
NGYALRF+I +VPVVG+KAAGVKAMNLK+ D + SAFI NT+S YLLT RGS�KR++ID

25 Sbjct: 661 QNGYALRFNIEEVPVVGAKAAGVKAMNLKEDDTLQSAFICNTSSFYLLTQRGSLKRVSID 720

Query: 720 VIPTTSRANRGLQVRELKSKPHRVFKAGFPVYLEDSSFEFDLFSSVSNHEGDTFVLEIMS 779
IP TSRA RGLQVRELK+KPHRVF AG V + F DLFS+ T L + S

30 Sbjct: 721 EIPATSRARGLQVRELKKNKPHRVFLAGSV--AEQGFVGDLFSTEVEENDQT--LIVQS 776

Query: 780 KTGKVDVDSQWSFSERTSNGSFVSDKISDEEVFSVKIK 819
G +Y+ L + SERTSNGSF+SD ISDEEVF +K

Sbjct: 777 NKGTTIESRLQDLNLSERTSNGSFISDTISDEEVFDAYLK 816

35 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2823> which encodes the amino acid sequence <SEQ ID 2824>. Analysis of this protein sequence reveals the following:

Possible site: 51
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -0.53 Transmembrane 376 - 392 (376 - 394)

40 ----- Final Results -----
bacterial membrane --- Certainty=0.1213(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

45

An alignment of the GAS and GBS proteins is shown below.

Identities = 633/819 (77%), Positives = 719/819 (87%)

50 Query: 1 MSNIQNMSLEDIMGERFGRYSKYIIQERALPDIRDGLKPQRRILYSMNKDGNTEKGFGR 60
MSNIQNMSLEDIMGERFGRYSKYIIQERALPDIRDGLKPQRRILYSMNKDGNTEKGR+R

Sbjct: 3 MSNIQNMSLEDIMGERFGRYSKYIIQERALPDIRDGLKPQRRILYSMNKDGNTEKGYR 62

Query: 61 KSAKSVGNVMGNFHPHGDSSIIDAMVRMSQDWNRETLIEMHGNNGSMDGDPAAAMRYTE 120
KSAKSVGN+MGNFHPHGDSSIIDAMVRMSQDWNRE L+EMHGNNGSMDGDP AAMRYTE

55 Sbjct: 63 KSAKSVGNIMGNFHPHGDSSIIDAMVRMSQDWNREILVEMHGNNGSMDGDPAAAMRYTE 122

Query: 121 ARLSEIAGYLLQDIDKNTVPFAWNFDDETEKEPTVLPAAFPNLLVNGATGISAGYATDIPP 180
ARLSEIAGYLLQDI+KNTV FAWNFDDETEKEPTVLPAAFPNLLVNG++GISAGYATDIPP

60 Sbjct: 123 ARLSEIAGYLLQDIEKNTVSAWNFDDETEKEPTVLPAAFPNLLVNGSSGISAGYATDIPP 182

Query: 181 HNLAEVIDAVVYMIDHPKAKLDKIMEFLPGDPFPTGAI IQGKDEIRKAYETGKGRVAVRS 240
HNL+EVIDAVVYMIDHPK L+KLMEFLPGDPFPTG IIQG DEI+KAYETGKGRV VRS

Sbjct: 183 HNLSEVIDAVVYMIDHPKASLEKLMFLPGDPFPTGGIIQGADEIKKAYETGKGRVVRS 242

65 Query: 241 RTAIETLKGKKQIIVTEIPYEVNKSVLVKRIDDVRRVNNKVPGLAEVRDESDDRLRIAI 300

-1020-

RT IE LKGGK+QIIVTEIPYEVNK+VLVK+IDDVRVNNKVPGI EVRDESDR GLRIAI
 Sbjct: 243 RTEIEELKGGKQIIVTEIPYEVNKAVLVKKIDDVRVNNKVPGIVEVRDESDRTGLRIAI 302

Query: 301 ELKKEADETIVLNYLFKYTDLQVNYNFMNVAIDYTPKQVGLSRILT/SYIAHRREIIAR 360
 ELKKEAD +LNYL KYTDLQVNYNFMNVAID +TP+QVGL +IL+SYI+HR++III R
 Sbjct: 303 ELKKEADSQTILNYLLKYTDLQVNYNFMNVAIDHFTPRQVGLQKILSSYSISRKDIIR 362

Query: 361 SKFDKEAEKRLHIVEGLIRVLSILDEVIALIRASENKADAKENLKVSYEFSEAQAEIV 420
 SKFDK AEKRLHIVEGLIRVLSILDE+IALIR+S+NKADAKENLKVSY+FSE QAEEIV
 Sbjct: 363 SKFDKAEKRLHIVEGLIRVLSILDEIIALIRSSDNKADAKENLKVSYDFSEEQAEIV 422

Query: 421 TLQLYRLTNIDIVTLREEEELRQQITMLKAIISDERTMYNVMKRELREVKKKFANTRRS 480
 TLQLYRLTNIDIVTL+ EE +LR IT L AII DE TMYNVMKRELREVKKKFAN R S
 Sbjct: 423 TLQLYRLTNIDIVTLQNEENDLRDLITLTSATIGDEATMYNVMKRELREVKKKFANPRLS 482

Query: 481 ELQELAEITIEIDTASLIEEDTYVSVTRGGYKRTSPRSFNASTVDELGKREDELIFVS 540
 ELQ ++ IEIDTASLI EE+T+VSVTRGGY+KRTSPRSFNAS+++E+GKR+DDELIFV
 Sbjct: 483 ELQAESQIIEIDTASLIAEEETVSVTRGGYKRTSPRSFNASSLEEVGKRDDDELIFVK 542

Query: 541 NAKTTQHLLMFTNLGNLAYRPVHELADIRWKDVGEHLSQNLVNFASNEEIIYAEIVDDFT 600
 AKTT+HLL+FT LGN+ YRP+HEL D+RWKD+GEHLSQ + NFA+ EEI+YA++V F
 Sbjct: 543 QAKTTEHLLFTTLGNVIYRPIHELTDLRWKDIGEHLSTISNFATEEEIYADIVTSFD 602

Query: 601 KETYFAVTSGLQIKRFERQEISPWRTYKSKTAKYAKLKSVEDYVVTVAPIQLEDVILVTY 660
 + Y AVT G IKRF+R+E+SPWRTYKSK+ KY KLK +D VVT++P+ +ED++LVT
 Sbjct: 603 QGLYVAVTQNGFIKRFDRKELSPWRTYKSKSTKYVKLDDKDRVVTLSPVIMEDLLLVTK 662

Query: 661 NGYALRFSINDVPVVGSKAAGVKAMNLDKRDHIVSAFIANTTSLYLLTHRGSLKRMADIV 720
 NGYALRFS +VP+ G K+AGVK +NLK+ D + SAF + S ++LT RGSLKRMAD
 Sbjct: 663 NGYALRFSQEVPIQGLKSAGVKGINLKNDDSLASAFVTSNSFFVLTQRGSLKRMADVD 722

Query: 721 IPTTSRANRGLQVLRRELKSKPHRVFKAGPVYLEDSSFEFDLFSSVSNHEGDTFVLEIMSK 780
 IP TSRANRGL VLRELK+KPHRVF AG V + S+ +FDLF+ + E + +LE++SK
 Sbjct: 723 IPQTSRANRGLLVLRRELKTKPHRVFLAGGVQSDTSAEQFDLFTDIPEEETNQQMLEVISK 782

Query: 781 TGKVVDVDSLQWSFSERTSNGSFVSDKISDEEVFSVKIK 819
 TG+ Y++ L S SER SNGSF+SD ISD+EV + +
 Sbjct: 783 TGQTYEIALETLSLSEISNGSFISDTISDQEVLVARTR 821

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 930

A DNA sequence (GBSx0986) was identified in *Sagalactiae* <SEQ ID 2825> which encodes the amino acid sequence <SEQ ID 2826>. Analysis of this protein sequence reveals the following:

Possible site: 49
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3369(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF64593 GB:AF169649 branched-chain aminotransferase IlvE
 [Lactococcus lactis]
 Identities = 259/340 (76%), Positives = 294/340 (86%)

Query: 1 MTVNLDWDNLGFAYRKLPPFRYISHFKDGKWDGKLTDDATLHISESSPALHYGQQAPEGL 60
 M +NLDW+NLGF+YR LPFRYI+ FKDGKW G+LT D LHISESSPALHYGQQ FEGL
 Sbjct: 1 MAINLDWENLGFYSYRNLPFRYIARFKDGKWSAGELTGDNQLHISESSPALHYGQQGFEG 60

Query: 61 KAYRTKDGSIQLFRPDQNAERLQRTADRLMPHVPTDKFIAAVKSVVRANEEFVPPYGTG 120

-1021-

KAYRTKDGSIQLFRRPDQNA RLQ+TA RL M V T+ FI AVK VV+AN++FVPPYGTG
 Sbjct: 61 KAYRTKDGSIQLFRRPDQNAARLQKTARRLCMAEVSTEMFIDAVKQVVKANKDFVPPYGTG 120
 Query: 121 ATLYIRPLLIGVGDIIGVKPAEEYIFTVFAMPVGSYFKGGLTPTNFIVSKEYDRAAPNGT 180
 5 ATLY+RPLLIGVGDIIGVKPA+EYIF VFAMPVGSYFKGGL P+ F++S+EYDRAAP GT
 Sbjct: 121 ATLYLRPLLIGVGDIIGVKPADEYIFKVFAMPVGSYFKGGLAPSKFVISREYDRAAPLGT 180
 Query: 181 GAAKVGGNYAASLLPGKYAHEKQFSDVIYLDPATHTKIEEVGAANFFGITKDNQFITPLS 240
 G AKVGGNYAASL A ++D IYLDP+THTKIEEVGAANFFGIT DN+FITPLS
 10 Sbjct: 181 GGAKVGGNYAASLQAEVGAASGYADAIYLDPSTHTKIEEVGAANFFGITADNEFITPLS 240
 Query: 241 PSILPSITKYSLLYLAKERFGMEAIEGDFVDELDKFTAGACGTAAVISPIGGIQNGDD 300
 PSILPSITKYSLLYLA+ R G++AIEG+V+ +L KF EAGACGTAA+ISPIG I +G+D
 15 Sbjct: 241 PSILPSITKYSLLYLAEHLGLKAIEGEVYAKDLGKFVEAGACGTAAIISPIGRIDDGED 300
 Query: 301 FHVFYSETEVGPATRKLYDELVGIQFGDVEAPEGWIVKVD 340
 ++F+SETEVGP ++LYDELVGIQFGDVEAPEGWI KVD
 Sbjct: 301 SYIFHSETEVGPVTKRLYDELVGIQFGDVEAPEGWIVKVD 340

20 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2827> which encodes the amino acid sequence <SEQ ID 2828>. Analysis of this protein sequence reveals the following:

Possible site: 61
 >>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1208(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

30 An alignment of the GAS and GBS proteins is shown below.

Identities = 280/340 (82%), Positives = 308/340 (90%)

Query: 1 MTVNLDWDNLGFAYRKLPFRYISHFKDGKWDGKLTDDATLHISESSPALHYGQQAPEGL 60
 MT+ +DWDNLGF Y KLPFRYIS++K+G+WD G+LT+DATLHISES+PALHYGQQAPEGL
 35 Sbjct: 16 MTIAIDWDNLGFYHKLPFRYISYYKNGQWDKGLTQEDATLHISESAPALHYGQQAPEGL 75
 Query: 61 KAYRTKDGSIQLFRRPDQNAERLQRTADRLIMPHVPTDKFIAAVKS VVRANEEFVPPYGTG 120
 KAYRTKDGSIQLFRRPD+NA RLQ TADRLIMP V T++FI A K VV+ANE+FVPPYGTG
 40 Sbjct: 76 KAYRTKDGSIQLFRRPDNAVRILQATADRLIMPQVSTEQFIDAAKQVVKANEDFVPPYGTG 135
 Query: 121 ATLYIRPLLIGVGDIIGVKPAEEYIFTVFAMPVGSYFKGGLTPTNFIVSKEYDRAAPNGT 180
 ATLY+RPLLIGVGDIIGVKPAEEYIFT+FAMPVG+YFKGGL PTNFIVS+ +DRAAP GT
 Sbjct: 136 ATLYLRPLLIGVGDIIGVKPAEEYIFTIFAMPVGNFYFKGGLAPTINFIVSEAFDRAAPYGT 195
 Query: 181 GAAKVGGNYAASLLPGKYAHEKQFSDVIYLDPATHTKIEEVGAANFFGITKDNQFITPLS 240
 GAAKVGGNYA SLLPGK A FSDVIYLDPATHTKIEEVGAANFFGIT +N+F+TPLS
 45 Sbjct: 196 GAAKVGGNYAGSLLPGKAASAGFSFSDVIYLDPATHTKIEEVGAANFFGITANNEFVTPLS 255
 Query: 241 PSILPSITKYSLLYLAKERFGMEAIEGDFVDELDKFTAGACGTAAVISPIGGIQNGDD 300
 PSILPSITKYSLL LA+ER GM IEGDV ++ELDKF EAGACGTAAVISPIGGIQ D+
 50 Sbjct: 256 PSILPSITKYSLLQLAEERLGMTVIEGDVPINELDKFVEAGACGTAAVISPIGGIQYKDN 315
 Query: 301 FHVFYSETEVGPATRKLYDELVGIQFGDVEAPEGWIVKVD 340
 HVFYSETEVGP TR+LYDELVGIQFGD+EAPEGWI KVD
 55 Sbjct: 316 LHVFYSETEVGPVTRRLYDELVGIQFGDIEAPEGWIVKVD 355

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 931

60 A DNA sequence (GBSx0987) was identified in *S.agalactiae* <SEQ ID 2829> which encodes the amino acid sequence <SEQ ID 2830>. Analysis of this protein sequence reveals the following:

-1022-

Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.3459(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9365> which encodes amino acid sequence <SEQ ID 9366>
10 was also identified. A further related GBS nucleic acid sequence <SEQ ID 10915> which encodes amino acid sequence <SEQ ID 10916> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2831> which encodes the amino acid sequence <SEQ ID 2832>. Analysis of this protein sequence reveals the following:

15 Possible site: 50
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

20 bacterial cytoplasm --- Certainty=0.3043(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

25 Identities = 22/36 (61%), Positives = 30/36 (83%)

Query: 4 IVSKKDKKIBIQISDAQVTVNGTKVDGYQLVMEKKL 39
 ++SKKDKKIBIQ+ D +V VN TK+DGYQL + K++
Sbjct: 1 VMSKKDKKIBIQLIDHKVMVNETKIDGYQLQIGKRV 36

30 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 932

A DNA sequence (GBSx0988) was identified in *S.agalactiae* <SEQ ID 2833> which encodes the amino acid sequence <SEQ ID 2834>. This protein is predicted to be glycyl-tRNA synthetase beta subunit (glyS).

35 Analysis of this protein sequence reveals the following:

Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

40 bacterial cytoplasm --- Certainty=0.1617(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

45 >GP:CAB73488 GB:AL139077 glycyl-tRNA synthetase beta chain
 [Campylobacter jejuni]

Identities = 33/90 (36%), Positives = 49/90 (53%), Gaps = 2/90 (2%)

50 Query: 3 RAFNLAEKVTHSVLVDSSLFENNQEKALYQAILSLELTEDMHDNLDKLFALSPIINDFFD 62
 R N+A K H V D SLF E LY+A + + L+ LFAL P I++FF+
Sbjct: 570 RLANIATKNPHKV--DESLFVQEAESKLYKAFQEKTKANSLQEKLENLFALKPFIDEFFN 627

Query: 63 NTMVMTDDEKMKQNLAILNSLVAKARTVA 92
 M+ +DEK+K NR A++ + A+ +A

-1023-

Sbjct: 628 QVMINADEKLNKNNRQALVYETIYAEFLKIA 657

There is also homology to SEQ ID 2836.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 933

A DNA sequence (GBSx0989) was identified in *S.galactiae* <SEQ ID 2837> which encodes the amino acid sequence <SEQ ID 2838>. Analysis of this protein sequence reveals the following:

Possible site: 30

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4825 (Affirmative) < succ>

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB13672 GB:Z99113 ynzC [Bacillus subtilis]

Identities = 41/72 (56%), Positives = 56/72 (76%)

Query: 5 KIARINELSKKKKT VGLTGEEKVEQAKLREEYIEGFRRSVRHHVEGIKLVDDEGNDVTPE 64

KIARINEL+ K K +T EEK EQ KLR+EY++GFR S+++ ++ +K++D EGNDVTPE

Sbjct: 6 KIARINELAAKAKAGVITEEBKAEQQKLRQEYLKGFSSMKNTLKS VKIIDPEGNDVTPE 65

Query: 65 KLRQVQREKGLH 76

KL++ QR LH

Sbjct: 66 KLRQVQREKGLH 77

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2839> which encodes the amino acid sequence <SEQ ID 2840>. Analysis of this protein sequence reveals the following:

Possible site: 30

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4303 (Affirmative) < succ>

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 79/85 (92%), Positives = 83/85 (96%)

Query: 1 MDPKKIARINELSKKKKT VGLTGEEKVEQAKLREEYIEGFRRSVRHHVEGIKLVDDEGND 60

MDPKKIARINEL+KKKKTVGLTG EKVEQAKLREEYIEG+RRSVRHH+EGIKLVD+EGND

Sbjct: 1 MDPKKIARINELAKKKKT VGLTGPEKVEQAKLREEYIEGYRRSVRHHIEGIKLVDDEGND 60

Query: 61 VTPEKLRQVQREKGLHGRSLDDPNS 85

VTPEKLRQVQREKGLHGRSLDDP S

Sbjct: 61 VTPEKLRQVQREKGLHGRSLDDPKS 85

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1024-

Example 934

A DNA sequence (GBSx0990) was identified in *S.galactiae* <SEQ ID 2841> which encodes the amino acid sequence <SEQ ID 2842>. Analysis of this protein sequence reveals the following:

```

Possible site: 20
5  >>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2343(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
10     bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAB69985 GB:U94355 glycerol kinase [Enterococcus casseliflavus]
Identities = 381/496 (76%), Positives = 439/496 (87%)

15 Query: 3  SEEKYIMAIHQGTSSRAIIFNKKGEKIASSQKEFPQIFPQAGWVEHNANQIWNQSVQSVI 62
      +E+ Y+MAIDQGTSSRAIIF++ G+KI SSQKEFPQ FP++GWVEHNAN+IWNQSVQSVI
      Sbjct: 2  AEKNYVMAIDQGTSSRAIIFDRNGKKIGSSQKEFPQYFPKSGWVEHNANEIWNQSVQSVI 61

20 Query: 63  AGAFIESSIKPGQIEAIGITNQRETTVVWDKKTGLPIYNAIVWQSRQIAPADQLKQEGH 122
      AGAFIES I+P I IGITNQRETTVVWDK TG PI NAIVWQSRQ++PIADQLK +GH
      Sbjct: 62  AGAFIESGIRPEAIGIGITNQRETTVVWDKKTGQPIANAIVWQSRQSSPIADQLKVDGH 121

25 Query: 123  TMNIHEKTGLVIDAYFSATKVRWILDHVPGAQERAEKGELEFGTIDTWLVWKLTDGLVHV 182
      T MIHEKTGLVIDAYFSATKVRW+LD++ GAQE+A+ GELLFGTID+WLWVWKLTDG VHV
      Sbjct: 122  TEMIHEKTGLVIDAYFSATKVRWLLDNIEGAQEKADNGELLEFGTIDSWLVWKLTDGQVHV 181

30 Query: 183  TDYSNAARTMLYNIKELKWDDEILELLNIPKAMLEPKVKSNEVYKTTTFFHYGGEVPI 242
      TDYSNA+RTMLYNI +L+WD EIL+LLNIP +MLPEVKSNSEVYK T +HFG EVPI+
      Sbjct: 182  TDYSNASRTMLYNIHLEWDQEILDLLNIPSSMLPEVKSNSEVYGHTRSYHFGSEVP 241

35 Query: 243  GMAGDQQAALFGQLAFEPGMVKNTYGTGSGFIIMNTGEEMQLSQNNLLTTIGYGINGKVHY 302
      GMAGDQQAALFGQ+AFE GM+KNTYGTG+FI+MNTGEE QLS N+LLTTIGYGINGKV+Y
      Sbjct: 242  GMAGDQQAALFGQMAFEKGMKNTYGTGAFIVMNTGEEPQLSDNDLLTTIGYGINGKVY 301

40 Query: 303  ALEGSIFVAGSAIQWLRDGLRMIETSSSESEGLAQSSSTSDDEVYVVPFTGLGAPYWDSEA 362
      ALEGSIF+AGSAIQWLRDGLRMIETS +SE LA + D+EVYVVPFTGLGAPYWDSEA
      Sbjct: 302  ALEGSIFVAGSAIQWLRDGLRMIETSPQSEELA KAKGDNEVYVVPFTGLGAPYWDSEA 361

45 Query: 363  RGSVFGTLRGTSKEDFVKATLQSIAYQVRDVIDTMQVDSGIDIQQLRVDGGAAMNLLMQ 422
      RG+VFGTLRGT+KEDFV+ATLQ++AYQ +DVIDTM+ DSGIDI L+VDGGA N+LLMQ
      Sbjct: 362  RGA VFGTLRGTTKEDFVRATLQAVAYQSKDVIDTMKKDSGIDIPLLKVDGGA AKNDLLMQ 421

50 Query: 423  FQADILGIDIDVQRAANLETTALGAAYLAGLVGFWKLDLDELKSMAREGQMFTPEMPAEERD 481
      FQADIL ID+ RA NLETTALGAA+LAGL+VG+W+ +DELK + GQ+F M ++
      Sbjct: 422  FQADILDIDVQRAANLETTALGAAYLAGLVGFWKLDLDELKSMAREGQMFTPEMPAEERD 481

Query: 483  KLYKGWRKAVKATQVF 498
      LY+GW++AV ATQ F
      Sbjct: 482  NLYEGWKQAVATQTF 497

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2843> which encodes the amino acid sequence <SEQ ID 2844>. Analysis of this protein sequence reveals the following:

```

Possible site: 19
55 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2282(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
60     bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

-1025-

Identities = 464/500 (92%), Positives = 484/500 (96%)

5 Query: 3 SEEKYIMAIIDQGTSSRAIIFNKKGEKIIASSQKEFPQIFPQAGWVEHNANQIWNVSQSVI 62
S+EKYIMAIIDQGTSSRAIIFN+KGEK++SSQKEFPQIFP AGWVEHNANQIWNVSQSVI
Sbjct: 2 SQEKYIMAIIDQGTSSRAIIFNQKGEKVSSSQKEFPQIFPHAGWVEHNANQIWNVSQSVI 61

10 Query: 63 AGAFIESSIKPGQIEAIGITNQRETTVVWDKKTGLPIYNAIVWQSRQTAPIADQLKQEGH 122
AGAFIESSIKP QIEAIGITNQRETTVVWDKKTG+PIYNAIVWQSRQTAPIA+QLKQ+GH
Sbjct: 62 AGAFIESSIKPSQIEAIGITNQRETTVVWDKKTGVPYNAIVWQSRQTAPIAEQLKQDGH 121

15 Query: 123 TNMIHEKTGLVIDAYFSATKVRWILDHVPGAQERAEGKELLFGTIDTWLVWKLTDGLVHV 182
T MIHEKTGLVIDAYFSATK+RWILDHVPGAQERAEGKELLFGTIDTWLVWKLTDG VHV
Sbjct: 122 TKMIHEKTGLVIDAYFSATKIRWILDHVPGAQERAEGKELLFGTIDTWLVWKLTDGAVHV 181

20 Query: 183 TDYSNAARTMLYNIKELKWDDEILELLNIPKAMLPVKSNSSEVYGKTTFFHYGGEVPIS 242
TDYSNAARTMLYNIK+L WDDEILELLNIPK MLPVKSNSSE+YGKT FFHYGGEVPIS
Sbjct: 182 TDYSNAARTMLYNIKDLTWDEILELLNIPKMDLPVKSNSSEIYGKTAAPFFHYGGEVPIS 241

25 Query: 243 GMAGDQQAALFGQLAFEPGMVKNTYGTGSFIIMNTGDEMQLSNNLLTTIGYGINGKVHY 302
GMAGDQQAALFGQLAFEPGMVKNTYGTGSFIIMNTG+EMQLS NNNLLTTIGYGINGKVHY
Sbjct: 242 GMAGDQQAALFGQLAFEPGMVKNTYGTGSFIIMNTGDEMQLSNNLLTTIGYGINGKVHY 301

30 Query: 303 ALEGSIFIAGSAIQWLRDGLRMIETSSESEGLAQSSSDDEVYVVPFTGLGAPYWDSNA 362
ALEGSIFIAGSAIQWLRDGL+MIETS ESE A +STDDEVYVVPFTGLGAPYWDSNA
Sbjct: 302 ALEGSIFIAGSAIQWLRDGLRMIETSPESQFALASTSDDEVYVVPFTGLGAPYWDSNA 361

35 Query: 363 RGSVFGLTRGTSKEDFVKATLQSIAYQVRDVIDTMQVDSGIDIQQLRVDGGAAMNNLLMQ 422
RGSVFGLTRGTSKEDFVKATLQSIAYQVRDVIDTMQVDSGIDIQQLRVDGGAAMNN+LMQ
Sbjct: 362 RGSVFGLTRGTSKEDFVKATLQSIAYQVRDVIDTMQVDSGIDIQQLRVDGGAAMNNMLMQ 421

40 Query: 423 FQADILGIDIDARAKNLETTALGAAPLAGLSVGYWBSMDELKELNATGQLFQATMNESRKE 482
FQADILGIDIDARAKNLETTALGAAPLAGL+VGYWE MD LKELNATGQLF+A+MNESRKE
Sbjct: 422 FQADILGIDIDARAKNLETTALGAAPLAGLAVGYWEDMDALKELNATGQLFKASMNESRKE 481

45 Query: 483 KLYKGWRKAVKATQVFAQED 502
KLYKGW++AVKATQVF QE+
Sbjct: 482 KLYKGWRKAVKATQVFTQEE 501

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
vaccines or diagnostics.

Example 935

A DNA sequence (GBSx0992) was identified in *S.agalactiae* <SEQ ID 2845> which encodes the amino acid sequence <SEQ ID 2846>. Analysis of this protein sequence reveals the following:

Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3146(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for
vaccines or diagnostics.

Example 936

A DNA sequence (GBSx0993) was identified in *Sagalactiae* <SEQ ID 2847> which encodes the amino acid sequence <SEQ ID 2848>. This protein is predicted to be alpha-glycerophosphate oxidase (glpD).

Analysis of this protein sequence reveals the following:

```

5   Possible site: 40
   >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -1.81    Transmembrane    20 - 36 ( 20 - 36)

10  ----- Final Results -----
      bacterial membrane --- Certainty=0.1723(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15  >GP:AAC34740 GB:U94770 alpha-glycerophosphate oxidase [Streptococcus pneumoniae]
    Identities = 464/608 (76%), Positives = 539/608 (88%)

Query: 1  MEFSRETRRLALQRMQDRTLDLLIIGGGITGAGVALQAAASGLDTGLIEMQDFAEGTSSR 60
      MEFS++TR L++++MQ+RTLDLLIIGGGITGAGVALQAAASGL+TGLIEMQDFAEGTSSR
20  Sbjct: 1  MEFSKKTRELSIKKMQERTLDLLIIGGGITGAGVALQAAASGLETGLIEMQDFAEGTSSR 60

Query: 61  STKL VHGGRLRYLKQFDVEVVSDTVSERAVVQIAPHIPKPDPMLLPVYDEPGSTFSMFR 120
      STKL VHGGRLRYLKQFDVEVVSDTVSERAVVQIAPHIPKPDPMLLPVYDE G+TFS+FR
25  Sbjct: 61  STKL VHGGRLRYLKQFDVEVVSDTVSERAVVQIAPHIPKPDPMLLPVYDEDGATFSLFR 120

Query: 121  KVAMDLYDL LAGVTNTPAANKVLSAEDVLKREPDLQKEGLGGGVYLD FRNNDARLV IEN 180
      KVAMDLYDL LAGV+NTP ANKVLS + VL+R+P+L+KEGL+GGGVYLD FRNNDARLV IEN
30  Sbjct: 121  KVAMDLYDL LAGVSNTP TANKVLSKDQVLERQPNLKK EGLVGGGVYLD FRNNDARLV IEN 180

Query: 181  IKRANRDGAYIASHVKAEDFLFDDNNQIIGVRARDLLTDQVIDIKARLVINTTGPWSDTV 240
      IKRAN+DGA IA+HVKA E FLFD++ +I GV ARDLLTDQV +IKARLVINTTGPWSD V
35  Sbjct: 181  IKRANQDGALIANHVKAEGFLFDESGKITGVVARDLLTDQVFEIKARLVINTTGPWSDKV 240

Query: 241  RNFSNECKQIHQLRPTKGVHLVVD RQKLNISQPVYDTGLNDGRMIFVLPREDKTYFGTT 300
      RN SN+G Q Q+RPTKGVHLVVD K+ +SQPVY DTGL DGRM+FVLPRE+KTYFGTT
40  Sbjct: 241  RNLSNKG TQFSQMRPTKGVHLVVDSSKIKV SQPVYDTGLGDGRMV FVLPRENKTYFGTT 300

Query: 301  DTDYHGDL EHPVTVKEDVDYLLNIVNKRFP EAE LTIDDI ESSWAGLRPLLSGNSASDYNG 360
      DTDY GDLEHP VT+EDVDYLL IVN RFPE+ +TIDDI ESSWAGLRPL++GNSASDYNG
45  Sbjct: 301  DTDYTGDL EHPKVTQEDVDYLLGIVNRRFPESNITIDDI ESSWAGLRPLIAGNSASDYNG 360

Query: 361  GNSGKLSDESFEELIDS VKDYIAHKNHREDVEKAISHVESSTSEKELDPSAVSRGSSFER 420
      GN+G +SDES F+ LI +V+ Y++ + REDVE A+S +ESSTSEK LDPSAVSRGSS +R
50  Sbjct: 361  GNNGTISDESFDNLIATVESYLSKEKTREDVESAVSKLESSTSEKHLDP SAVSRGSS LDR 420

Query: 421  DDNGLLTLAGGKITDYRKMAEGAMETIINILDK EYNRKFKLINSKTYPVSGGEINPSNVD 480
      DDNGLLTLAGGKITDYRKMAEGAME +++IL E++R FKLINSKTYPVSGGE+NP+NVD
55  Sbjct: 421  DDNGLLTLAGGKITDYRKMAEGAMERVVDILKAEFDRSFKLINSKTYPVSGGELNPANVD 480

Query: 481  SEIEAQAQLGTL SGLSIEDARYIANLYGSNAPKLFALTRQITEAGLSLVETLSLHYAMD 540
      SEIEA+AQLG GL ++A Y+ANLYGSNAPK+FAL + +A GLSL +TSLHYAM
60  Sbjct: 481  SEIEAQAQLGVSRGLDSKEAHYIANLYGSNAPKVFALAHSL EQAPGLSLADTSLHYAMR 540

Query: 541  YEMALSPTDFFLRRTNHMLFMRDNLDSL IQFVIDEMAKHYQWSDQDKTFYEEELHETLKD 600
      E+ALSP DF LRRTNHMLFMRD+LDS+++PV+DEM + Y W+++K Y ++ L +
65  Sbjct: 541  NELALSPVDFLLRRTNHMLFMRDNLDSIVEPVLD E MGRFYDWTEEEKATYRADVEAALAN 600

Query: 601  NDLAALKD 608
      NDLA LK+
66  Sbjct: 601  NDLAELKN 608

```

There is also homology to SEQ ID 128.

-1027-

SEQ ID 2848 (GBS93) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 7 (lane 7; MW 70.6kDa).

GBS93-His was purified as shown in Figure 192, lane 4.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 937

A DNA sequence (GBSx0994) was identified in *S.agalactiae* <SEQ ID 2849> which encodes the amino acid sequence <SEQ ID 2850>. Analysis of this protein sequence reveals the following:

```

Possible site: 23
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.0965(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 938

A DNA sequence (GBSx0995) was identified in *S.agalactiae* <SEQ ID 2851> which encodes the amino acid sequence <SEQ ID 2852>. This protein is predicted to be glycerol uptake facilitator protein (glpF). Analysis of this protein sequence reveals the following:

```

Possible site: 55
>>> Seems to have a cleavable N-term signal seq.
      INTEGRAL    Likelihood = -7.43    Transmembrane  220 - 236 ( 216 - 236)
      INTEGRAL    Likelihood = -6.48    Transmembrane  139 - 155 ( 136 - 158)
      INTEGRAL    Likelihood = -3.88    Transmembrane   87 - 103 (  83 - 107)
      INTEGRAL    Likelihood = -3.03    Transmembrane  164 - 180 ( 162 - 183)

----- Final Results -----
      bacterial membrane --- Certainty=0.3972(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 8689> which encodes amino acid sequence <SEQ ID 8690> was also identified. Analysis of this protein sequence reveals the following:

```

Lipop: Possible site: -1    Crend: 5
SRCEFLG: 0
McG: Length of UR:  21
      Peak Value of UR:  2.51
      Net Charge of CR: -2
McG: Discrim Score:  4.43
GvH: Signal Score (-7.5): -0.139999
      Possible site: 50
>>> Seems to have a cleavable N-term signal seq.
Amino Acid Composition: calculated from 51
ALOM program  count: 4 value: -7.43 threshold: 0.0
      INTEGRAL    Likelihood = -7.43    Transmembrane  215 - 231 ( 211 - 231)

```

-1028-

INTEGRAL Likelihood = -6.48 Transmembrane 134 - 150 (131 - 153)
 INTEGRAL Likelihood = -3.88 Transmembrane 82 - 98 (78 - 102)
 INTEGRAL Likelihood = -3.03 Transmembrane 159 - 175 (157 - 178)
 PERIPHERAL Likelihood = 4.98 65
 modified ALOM score: 1.99
 icml HYPID: 7 CFP: 0.397

*** Reasoning Step: 3

----- Final Results -----

bacterial membrane --- Certainty=0.3972(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAA91618 GB:U12567 glycerol uptake facilitator [Streptococcus pneumoniae]
 Identities = 150/230 (65%), Positives = 194/230 (84%), Gaps = 1/230 (0%)

Query: 7 DIFGEFLGTALLVLLGNGVVAGVVLPTKKNHNSGWIVITFGWGLAVAIAALVSGNISPAH 66
 ++FGEFLGT +L+LLGNGVVAGVVLPTK+++SGWIVIT G+AVA+A VSG +SPA
 Sbjct: 4 ELFGEFLGTLLILLGNGVVAGVVLPTKSNSSGWIVITMV-GIAVAVAVFVSGKLSPAH 62

Query: 67 LNPVSLAFAIKGDLAWGTAILYMIAQIIGAMLSLLVYLQFRPHYEAENRADILGTFA 126
 LNPV++ A+KG L W + + Y++AQ GAMLG +LV+LQF+PHYEA EN +IL TF+
 Sbjct: 63 LNPVITIGVALKGGPLWASVLPYILAQFAGAMLGQILVWLQFKPHYEAENAGNILATFS 122

Query: 127 TGPALKDNFSNFLSEVLGTLVLVLTIFAIGKYNMPPGVGTMSVGMVLVVGIGLSLGGTTGY 186
 TGPA+KD SN +SE+LGT VLVLTIFA+G Y+ G+GT +VG L+VGIGLSLGGTTGY
 Sbjct: 123 TGPAIKDTSNLSLISEILGTFVLVLTIFALGLYDFQAGIGTFAVGTLLIVGIGLSLGGTTGY 182

Query: 187 AINPARDFGPRLLHALLPKMKNGSDWTYSWIPIVGPVMGAILAALIFAM 236
 A+NPARD GPR++H++LP+ NKGD DW+Y+WIP+VGP++GA LA L+F++
 Sbjct: 183 ALNPARDLGERIMHSTLPIPNKGDGDSYAWIPVVGPFVIGAALAVLVFSL 232

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2853> which encodes the amino acid sequence <SEQ ID 2854>. Analysis of this protein sequence reveals the following:

Possible site: 50

>>> Seems to have a cleavable N-term signal seq.

INTEGRAL Likelihood = -9.13 Transmembrane 213 - 229 (209 - 232)
 INTEGRAL Likelihood = -5.52 Transmembrane 137 - 153 (132 - 157)
 INTEGRAL Likelihood = -4.35 Transmembrane 159 - 175 (155 - 178)
 INTEGRAL Likelihood = -1.17 Transmembrane 85 - 101 (85 - 101)

----- Final Results -----

bacterial membrane --- Certainty=0.4652(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAA91618 GB:U12567 glycerol uptake facilitator [Streptococcus pneumoniae]
 Identities = 159/230 (69%), Positives = 196/230 (85%), Gaps = 1/230 (0%)

Query: 2 DIFGEFLGTALLVLLGNGVVAGVVLPTKTHASGWIVITATGWGIAVAVAVFISGKVAPAH 61
 ++FGEFLGT +L+LLGNGVVAGVVLPTK+++SGWIVI T GIAVAVAVF+SGK++PAH
 Sbjct: 4 ELFGEFLGTLLILLGNGVVAGVVLPTKSNSSGWIVI-TMVGIAVAVAVFVSGKLSPAH 62

Query: 62 LNPVSLAFAMSGTIAWSTAIAYSLAQLLGAMVGSTLVFLQFRPHYLAESQADILGTFA 121
 LNPV++ A+ G + W++ + Y LAQ GAM+G LV+LQF+PHY A E+ +IL TF+
 Sbjct: 63 LNPVITIGVALKGGPLWASVLPYILAQFAGAMLGQILVWLQFKPHYEAENAGNILATFS 122

Query: 122 TGPAIRDTSSNLLSEIFGTFLVLMGLAFLGYDMPAGLGTLCVGTLLVIGIGLSLGGTTGY 181
 TGPAI+DT SNL+SEI GTFVL+L I A GLYD AG+GT VGTLL+GIGLSLGGTTGY
 Sbjct: 123 TGPAIKDTSNLSLISEILGTFVLVLTIFALGLYDFQAGIGTFAVGTLLIVGIGLSLGGTTGY 182

-1029-

Query: 182 AINPARDLGPRLVHAILPLNNKGDSDWSYAWIPVVGPIIGAVLAVLLFQV 231
 A+NPARDLGPR++H+ILP+ NKGD DWSYAWIPVVGPI+IGA LAVL+F +
 Sbjct: 183 ALNPARDLGPRIMHSILPINKGDSDWSYAWIPVVGPIIGAAVLAVLVSFL 232

5 An alignment of the GAS and GBS proteins is shown below.

Identities = 169/232 (72%), Positives = 202/232 (86%)

Query: 6 MDIFGEFLGTALLVLLGNGVVAGVVLPTKKNHNSGWIVITFGWGLAVAIAALVSGNISPA 65
 MDIFGEFLGTALLVLLGNGVVAGVVLPTK H SGWIVI GWG+AVA+A +SG ++PA
 10 Sbjct: 1 MDIFGEFLGTALLVLLGNGVVAGVVLPTKTHASGWIVITATGWGIAVAVAVFISGKVAPA 60

Query: 66 HLNPAVSLAFAIKGDLAWGTAILYMQIIGAMLGSLVLYLQFRPHYEAENRADILGTF 125
 HLNPAVSLAFA+ G +AW TAI Y +AQ++GAM+GS LV+LQFRPHY AAE++ADILGTF
 15 Sbjct: 61 HLNPAVSLAFAMSGTIAWSTAIYSLAQLLGAMVGSTLVFLQFRPHYLAESQADILGTF 120

Query: 126 ATGPALKDNFSNFLSEVLGTLVLVLTIFAIGKYNMPPGVGTMSVGMVVGIGLSLGGTTG 185
 ATGPA++D SN LSE+ GT VL+L I A G Y+MP G+GT+ VG LV+GIGLSLGGTTG
 20 Sbjct: 121 ATGPAIRDTSSNLLSEIFGTFLMLGILAFGLYDMPAGLGLTLCVGTLVIGIGLSLGGTTG 180

Query: 186 YAINPARDFGPRLHLHALLPMKNKGDSDWTYSWIPVGPVVGAILAALIFAMM 237
 YAINPARD GPRL+HA+LP+ NKGDSDW+Y+WIP+VGP++GA+LA L+F +M
 25 Sbjct: 181 YAINPARDLGPRLVHAILPLNNKGDSDWSYAWIPVVGPIIGAVLAVLLFQVM 232

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 25 vaccines or diagnostics.

Example 939

A DNA sequence (GBSx0996) was identified in *S.galactiae* <SEQ ID 2855> which encodes the amino acid sequence <SEQ ID 2856>. This protein is predicted to be NADH oxidase. Analysis of this protein sequence reveals the following:

30 Possible site: 23
 >>> Seems to have a cleavable N-term signal seq.
 INTEGRAL Likelihood = -2.87 Transmembrane 152 - 168 (152 - 168)

35 ----- Final Results -----
 bacterial membrane --- Certainty=0.2147(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9523> which encodes amino acid sequence <SEQ ID 9524>
 40 was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA48728 GB:X68847 NADH oxidase [Enterococcus faecalis]
 Identities = 105/423 (24%), Positives = 197/423 (45%), Gaps = 15/423 (3%)

45 Query: 10 IVILGASFAGMTCAQKLRLNPNWDIVLIDKEIHDPYVPGNLNWWYRHEISGLNQAMWQT 69
 +V++G + AG + + + +P ++ + ++ + ++ G+ Y + +
 Sbjct: 3 VVVVGCTHAGTSAVKSILANHPAEVTVYERNDNISPLSCGIALYVGGVVKNAADLFYSN 62

50 Query: 70 EEEQRLQNIRCLFGLKVEKINKEDR-----ELMLSDGSSVYYDQLICAMGSAESTYIDG 124
 EE VE+IN +D+ L +V YD+L+ GS I G
 Sbjct: 63 PEELASLGATVKMEHNVEEINVDKTVTAKNLQTGATETVSVDKLVMTTGSWPIIPPIG 122

55 Query: 125 ADAQGVLTTKTYATSONAKQVLDKSHKVAVVGAGIIGLDIAYSLHESGKAVTLLAQERP 184
 DA+ +L K Y+ + + + +V VVG G IG+++ + ESGK VTL++ +R
 Sbjct: 123 IDAENILLCKNYSQANVIEKAKDAKRVVVGGGYIGIELVEAFVESGKQVTLVDGLDRI 182

Query: 185 DFRHTDPDMSLPLLDAMAESKLHFFQNKVEKITVTREKLCIRTLTGDTFTVDVAVILAV 244
 ++ D + L + + ++ + V++ + K+ F D VI+ V

-1030-

Sbjct: 183 LNKYLDKPFPTDVLEKELVDRGVNLAGENVQQFVADEQKQKAVKITPSQEFEDMVIMCV 242
 Query: 245 NFRPDSRLITGLVDLSVDNSVVNDYFQTSDPNIYAIGDLIWSYFKGLNSAYYMPILINQA 304
 FRP++ LL VD+ + ++ VN+Y QTS+P+I+A GD ++ + Y+PL A
 Sbjct: 243 GFRPNTLLKDKVDMLPNGAIEVNEYMQTSNPDIFAAGDSAVVHYNPSQTKNYIPLAINA 302
 Query: 305 IRSAQMLAYHLSGHAVPKLKITRATGSKHFCYYRANIGLT-----ELEAGFYEDTV 355
 +R ++ +L+ + +G FG+ + G+T ++EA +ED
 Sbjct: 303 VRQGMVLVGRNLTEQKLAYRGQTGTSGLYLFQWKIGSTGVTKESAKINGLDVEATVFEDNY 362
 Query: 356 SVTYFPKEQYDLRIKLIANQKTGHLGLGAQLISKENCLATANQLVQAISCDMTDFDLAFQD 415
 + P + L ++L+ + T ++G QL+SK + +AN L A+ MT DLA D
 Sbjct: 363 RPEFMPTEKVL-MELVYEKGTQRIVGGQLMSKYDITQSANTLSLAVQNKMTVEDLAISD 421
 Query: 416 FIY 418
 F +
 Sbjct: 422 FFF 424

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2857> which encodes the amino acid
 sequence <SEQ ID 2858>. Analysis of this protein sequence reveals the following:

Possible site: 16
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -3.35 Transmembrane 155 - 171 (155 - 173)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.2338(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
 RGD motif: 54-56

The protein has homology with the following sequences in the databases:

>GP:CAA44611 GB:X62755 NADH peroxidase [Enterococcus faecalis]
 Identities = 111/428 (25%), Positives = 202/428 (46%), Gaps = 24/428 (5%)
 Query: 10 VIGASFAGLAFVDKYKDLNPDSSQIILIDKESCPNYIPNGINQLFRGDIQDLSDAMWGRAC 69
 V+G+S G V++ +L+PD++I +K +++ G+ G ++D++ R
 Sbjct: 5 VLGSSHGGYEAEEELNLHPDAEIQWYKGFISFLSCGMQLYLEGKVKDVNSV---RYM 61
 Query: 70 LAAQIESN--HRFIQAEVLAIEAPSNTLLKDS-QGRVFEEGYETLVCAMGASPOSHYIE 126
 ++ES + F E+ AI+ + + +KD G E Y+ L+ + GA P I
 Sbjct: 62 TGEKMSRGNVNFSENTEITAIQPKHQVTVKDLVSGEERVENYDKLIISPGAVPFELDIP 121
 Query: 127 TSQTNKVLVTKYYESQASLKLEASQE-----VLVIGAGLIGLDLAYSLSLQKRVKLI 181
 + + + + Q ++KL + + + V+VIG+G IG++ A + + GK+V +I
 Sbjct: 122 GKDLNIIYLMR---GRQWAIKQKQTVDPVNNVVVIGSGYIGIEAAEFAKAGKQVTVI 178
 Query: 182 EAAERPDFYQTDALIAPVMAEMSTHHVTFINNKRVTAIHEIEGKVVAHTEQGDTFQGDIL 241
 + +RP D E + EM +++T + V +E +G+V + + DL
 Sbjct: 179 DILDRPLGVLDKEFTDVLTEEMEANNITITATGETVER-YEGDGRVQKVVDKNAYDADL 237
 Query: 242 AILAINFRPNTHLLQGQVACALDKTILVNENLQTSQANIYAIGDMVSLHFGILGMDYYTP 301
 ++A+ RPNT L+G + + I +E ++TS+ +++A+GD + + +
 Sbjct: 238 VVVAVGVPRNTAWLKGTLELHPNGLIKTDEYMRTPSEPDVFAVGDATLIKYNPADTEVNIA 297
 Query: 302 LINQAMKTGQALALHLGYPPIPLQTVK-VLGSSHFDYYRASVCVTE-----EEAEYI 353
 L A K G+ +L P+ P V+ G + FDY AS G+ E +E +
 Sbjct: 298 LATNARKQGRFAVKNL-EPVKFPFGVQGSGLAVFDYKFASTGINEVMAQKLGKETKAV 356
 Query: 354 MDTCSYLYQNGDSKNLFWLKLIAKRTDGLILGAQLLSKTNALVIANQLGQALALKVTDAD 413
 YL K W KL+ ++GAQL+SK + N + A+ K+T D
 Sbjct: 357 TVVEDYLMDFNPDKQKAWFKLVYDPETTQILGAQLMSKADLTANINAISLAIQAKMTIED 416
 Query: 414 LAFQDFLF 421

-1031-

LA+ DF F
Sbjct: 417 LAYADFFF 424

An alignment of the GAS and GBS proteins is shown below.

5 Identities = 192/440 (43%), Positives = 276/440 (62%), Gaps = 7/440 (1%)

Query: 8 KVIVILGASFAGMTCAQKLRQLNPNWDIVLIDKEIHDPDYVPNGLNWWYRHEISGLNQAMW 67
K I ++GASFAG+ K + LNP+ I+LIDKE P+Y+PNG+N +R +I L+ AMW
10 Sbjct: 6 KTIHVIGASFAGLAFVDKYKDLNPDSSQIILIDKESCPNYIPNGINQLFRGDIQDLSAMW 65

Query: 68 -QTREEQRLQNIRCLFGLKVEKINKEDRELMLSDGSSVY----YDQLICAMGSAESTYI 122
+ +++++ +V I L+L D Y+ L+CAMG+ +S YI
15 Sbjct: 66 GRACLAQAQIESNHRFIQAEVLAIEAPSNTLLKDSQGRVFEEGYETLVCAMGASPOSHYI 125

Query: 123 DGADAQGVLTTTKTYATSQNAKQVLDKSHKAVVAGAGIIGLDIAYSLHESGKAVTLLEAQE 182
+ + VL TK Y SQ + +++++ S +V V+GAG+IGLD+AYSL GK V L+EA E
20 Sbjct: 126 ETSQTNKVLVTKYEESQASLKLEASQEVLVIGAGLIGLDLAYSLSLQKRVKLEAAE 185

Query: 183 RPDFRHTDPMDSLPDLLDAMAESKLHFFQNKVEKITVTREEKLCRLTLTGDTFTVDVAIL 242
RPDF TD ++ P++ M+ + F N++V I E K+ T GDTF D IL
25 Sbjct: 186 RPDFVQTDABELIAPVMAEMSTHHVTFINNKRVTAIHEI-EGKVVAHQEQGDTFQGDILAIL 244

Query: 243 AVNFRPDSRLLTGLVDLSVDNSVVVNDYFQTSDFPNYAIGDLIWSYFKGLNSAYYMPILIN 302
A+NFRP++ LL G V ++D +++VN+ QTS NIYAIGD++ +F L YY PLIN
30 Sbjct: 245 AINFRPNTHLLQGQVACALDKTILVNENLQTSQANIYAIGDMVSLHFGILGMDYYTPLIN 304

Query: 303 QAIRSAQMLAYHLSGHAVPKLKITRATGSKHFGYYRANIGLTEAGFYEDTVSVTYFPK 362
QA+++ Q LA HL+G+ +P L+ + GS HF YYRA++G+TE EA Y DT S Y
35 Sbjct: 305 QAMKTGQALALHLAGYPIPLQTVKVLGSSSHFDYYRASVGVTEEEAEYMDTCSYLYQNG 364

Query: 363 EQYDL-RIKLIANQKTGHLIGAQLISKENCLATANQLVQAISCDMTDFDLAFQDFIYTAR 421
+ +L +KLIA + G L+GAQL+SK N L ANQL QA++ +TD DLAFQDF++
40 Sbjct: 365 DSKNLFWLKLIARKTDGILIGAQLLSKTNALVIANQLGQALALKVTDADLAFQDFLFLQG 424

Query: 422 ESEMAYMLHQAAILNLYEKRI 441
S++AY LH+A + L+EKR+
45 Sbjct: 425 HSDLAYHLHEACKLFEKRL 444

There is also homology to SEQ IDs 1820, 1876, 4666.

40 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 940

A DNA sequence (GBSx0998) was identified in *S.galactiae* <SEQ ID 2859> which encodes the amino acid sequence <SEQ ID 2860>. Analysis of this protein sequence reveals the following:

45 Possible site: 31
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
50 bacterial cytoplasm --- Certainty=0.2980(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

55 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 941

A DNA sequence (GBSx0999) was identified in *S.agalactiae* <SEQ ID 2861> which encodes the amino acid sequence <SEQ ID 2862>. Analysis of this protein sequence reveals the following:

Possible site: 23
 5 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3548(Affirmative) < succ>
 10 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for
 15 vaccines or diagnostics.

Example 942

A DNA sequence (GBSx1000) was identified in *S.agalactiae* <SEQ ID 2863> which encodes the amino acid sequence <SEQ ID 2864>. Analysis of this protein sequence reveals the following:

Possible site: 29
 20 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1685(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 25 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9525> which encodes amino acid sequence <SEQ ID 9526> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

30 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2865> which encodes the amino acid sequence <SEQ ID 2866>. Analysis of this protein sequence reveals the following:

Possible site: 22
 >>> Seems to have no N-terminal signal sequence

35 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3125(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

40 An alignment of the GAS and GBS proteins is shown below.

Identities = 179/476 (37%), Positives = 279/476 (58%), Gaps = 5/476 (1%)

Query: 1 MRIEALMEKERRVQYRLLSFLRGSPQAIALKLALLETGLSRATFLKYINNLSYFEQEKV 60
 M+IE IM+KERR QYRL L + + + LK + + LS+ T LKYI+NLN ++ +
 45 Sbjct: 21 MKIEDLMDKERRAQYRLIVTLYHAKETLRKDLMLRLSNLSKVTLKYIDNLNHLCREQGL 80

Query: 61 NCRIVYKDKLFLEEDYNLSNQEV LKALMKDSIKYITILISLFNQRQFTIVGLSQELMVSE 120
 C+++ KD L L+E+ +++++ L+K+S+ Y IL ++ F I LS ELMVSE
 Sbjct: 81 ACQLLEKDSLSLKENGQFHWEDLVALLKESVAYQILTYMYCHEHFNITNLSVELMVSE 140

50 Query: 121 ATLNRHLAHLNELLAEFDIAISQKGIGDELQWRYFYFELFKQLWSYDKCQNMIKKLDLD 180
 ATLNR LAHLN+LL+EFD+A+SQG+Q+G ELQWRYFY+ELF+ + ++ +LD

-1033-

Sbjct: 141 ATLNRQLAHLNQLLSEFDLALSQGRQLGSELQWRYFYFELFRHTLTTRQGIDALVNQLDAS 200

Query: 181 SLILLIERLAQHTLTREAHQNLGLWFSICHRRLLAMEKISDNLKPIVKHYQCNAFYKRLD 240
 L LIERL +L+ EA + L +W +I R+ + +D+ N F+KRL+

5 Sbjct: 201 HLATLIERLIGQSLSAEALQLLIWLAI SQARMSFQKSYNDHFLRDSDFMTSNIFFKRL 260

Query: 241 AALVLYMSRFALEYREGEVLATFAFLHSQNILPINTMEYIMGFGGPIIDCVTETIIYFKK 300
 + L+ Y+ R+ALE+ E + F FLH+ +LPI +M+Y +GFGGPI D ++E + KK

10 Sbjct: 261 SMLLHYLRRYALEFDFAFEAKSLFVFLHAYPLLP IASMKYSLGFGGPIADHISEALWLLKK 320

Query: 301 ESILADETSDQVIYQLGQLYSHYYFFKGHILVEQPDLEQTYRLIDHNM RDKLHHISKKII 360
 ++ +T +++IY LG +S YPFKG IL + + + Y+L+ + R L I ++

15 Sbjct: 321 AHVIIHQTKKEIYGLGIFFSKAYFFKGAILSQPTINSQYLYQLVGEDKRALLRVIINHLV 380

Query: 361 ANVNRIRPLTEDGCSLLTLHLLELLIFSKNSQKMPFRIGLDMTGNAVEQSLL EYRIRQHF 420
 +++ D L+ +L LLIFS P +GL + N VE ++ E IR+H

20 Sbjct: 381 LQMDQ----ETDFSQQLSDDILALLIFSIERHHEPLLVLGLALGQNKVEAAIAELAIRRHL 436

Query: 421 SGNNSIQVEPYDEGKGFDMVIYQSHSRPYKALITYCLNKGASERELQEIDSLIYD 475
 Q+ PYD K +D ++ YQ+ P + Y L + +S EL +++ + D

Sbjct: 437 GHRDFQLMPYDHQKVYDCLITYQTVC LPRQDLFPYRLKQYSSPYELTALEAFLKD 492

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

25 Example 943

A DNA sequence (GBSx1001) was identified in *Sagalactiae* <SEQ ID 2867> which encodes the amino acid sequence <SEQ ID 2868>. This protein is predicted to be transketolase (tktA-1). Analysis of this protein sequence reveals the following:

Possible site: 27

30 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2084(Affirmative) < succ>

35 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9527> which encodes amino acid sequence <SEQ ID 9528> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

40 >GP:BAB06071 GB:AP001515 transketolase [Bacillus halodurans]

Identities = 403/661 (60%), Positives = 520/661 (77%), Gaps = 8/661 (1%)

Query: 6 IDQLAVNTVRTLSIDAIAQANS GHPLPMGAAPMAYVLWNKFLNVNPKTSRNWTRDRFV 65
 ++QLAVNT+RTLSID+++ ANSGHPG+PMGAAPMA+ LW KF+N NP + +W NRDRFV

45 Sbjct: 5 VEQLAVNTIRTLSIDSVEKANS GHGMPGMPGAAPMAFCLWTKFMNHNP-ANPDWVNRDRFV 63

Query: 66 LSAGHGSALLYSLLHLAGYDLSIDDLKQFRQWGSKTPGHPEVNHTDGVEATTGPLGQGIA 125
 LSAGHGS LLYSLLHL GYDLS+++L+ FRQWGSKTPGHPE HT GVEATTGPLGQG+A

50 Sbjct: 64 LSAGHGSMLLYSLLHLTG YDLSLEELQNFRQWGSKTPGHPEYGHTPGVEATTGPLGQGVA 123

Query: 126 NAVGMAMAEAHAAKFNKPGFDLV DHYTTYTLHGDGCLMEGVSQEASLAGHLKLGKLVLL 185
 AVGMAMAE HLAA +N+ G+++VDHYTTYT+ GDG LMEGV S EAASLAGHLKLG+++LL

Sbjct: 124 MAVGMAMAEERHLAATYNRDGYNIVDHYTTYTCGDGDLMEGVSAEAASLAGHLKLG RMILL 183

55 Query: 186 YDSNDISLDGPTSQSFTEDVKGRFESYGWQHILVKDGNLDEAIAAAIEA AKAETDKPTII 245
 YDSNDISLDG SF+E V+ RF++YGW + V+DGN+L+ IA AIE AKA+ ++P++I

Sbjct: 184 YDSNDISLDGDLHHSFSESVEDRFRKAYGWHVVRVEDGNNLDEIAK AIEAKAD-ERPSLI 242

Query: 246 EVKTTIIGFGAEKQGTSSV-HGAPLGARGITFAKAYVWEYP-DFTVPAEVADRFASDLQA 303

-1034-

EVKT IGFG+ +G SV HGAPLGA+ + K+AY W Y +F +P EVA + ++
 Sbjct: 243 EVKTTIGFGSPNKGKSVSHGAPLGADEVKLTKEAYEWTYENEFHIPEVA-AYYEQVKK 301

Query: 304 RGAKEEAANDLFAKYEVEYPELATEYKEAFAG---QAETVELKAHDLGSSVASRVSSQQ 360
 +GA+ EE+WN+LFA+Y+ YPELA++++ A G + ++++G SVA+R SS +
 Sbjct: 302 QGAKEESWNELFAQYKAYPELASQFELAVHGDLPFCWDAVAPSYEVGKSVATRSSSGE 361

Query: 361 AIQQLSTQLPNLWGGADLSASNNIMVAAETDFQASNYAGRNIWFGVREFAMAAAMNGIA 420
 A+ + +P L+GGSADL++SN T++ E +F +Y+GRN+WFGVREFAM AAMNG+A
 Sbjct: 362 ALNAFAKTVPQLFGGSADLASSNKTILIKGEANFSRDDYSGRNVWFGVREFAMGAAMNGMA 421

Query: 421 LHGGTRVYGGTFFVFSNYLLPAVRMAALQNLPTVYVMTDHSIAGVEDGPTHEPIEQLASV 480
 LHGG +V+G TFFVFS+YL PA+R+AAL LP +YV THDSIAGVEDGPTHEP+EQLAS+
 Sbjct: 422 LHGGLKVFSGTFFVFSNYLLPAVRMAALQNLPTVYVMTDHSIAGVEDGPTHEPVEQLASL 481

Query: 481 RSMPLNVIRPADGNETNAAWQRAVSETDRPTMLVLTRQNLPLVLEGTSELAQEGVKNKGAY 540
 R+MP L+VIRPADGNE+ AAW+ A+ D+PT LVL+RQNLPL LEG + A +GV+KGAY
 Sbjct: 482 RAMPGLSVIRPADGNEVAWKLALSKDQPTALVLSRQNLPTLEGAVDRAYDGVSKGAY 541

Query: 541 ILSEAKGELDGIIATGSEVKLALDTQDKLESEGIHVRVVSMPAQNIFDEQEASYQEQLV 600
 +L+ A G D +++A+GSEV LA++ ++ LE EGIH VVSMP+ + F+ Q A Y+E+VL
 Sbjct: 542 VLPANGSADLLLLASGSEVSLAVNAKEALEKEGIHAHVSMPSWDRFEAQSAEYKEEVL 601

Query: 601 PSAVTKRLAIEAGSSFGWGKYVGLNGLTLTIDTWGASAPGNRIFEEYGFVENAVSLYKEL 661
 PS VT RLAIIE GSS GW KYVG G + ID +GASAPG RI EE+GFTV++ V+ K L
 Sbjct: 602 PSDVTARLAIEMGSSSLGWAKYVGNQGDVVAIDRFASAPGERIMEEFGFTVQHVVVARAKAL 662

There is also homology to SEQ ID 520.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 944

A DNA sequence (GBSx1002) was identified in *S. agalactiae* <SEQ ID 2869> which encodes the amino acid sequence <SEQ ID 2870>. Analysis of this protein sequence reveals the following:

Possible site: 39
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4477(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9529> which encodes amino acid sequence <SEQ ID 9530> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 2871> which encodes the amino acid sequence <SEQ ID 2872>. Analysis of this protein sequence reveals the following:

Possible site: 42
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4581(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 27/79 (34%), Positives = 45/79 (56%)

-1035-

Query: 3 MKKECRDFYRQIQHTYNDISVREDAVLSSILLSASNGLIKTSVPRVAYELTQQLENNEI 62
 M+K+ + Y I+ Y+ RE+ LS +LL+ASN LIK S+ VAY+L Q ++N +
 Sbjet: 1 MEKKRQRLYDVIRQAYDYPENRENVALSQLLLAASNRLIKHSNPLLVAQLNQDQVDNYLL 60

Query: 63 EKSFESESLATVKELKKSAAK 81
 + ++ K+S +K
 Sbjet: 61 DNDILLPKSLCRFKQSLEK 79

- 10 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 945

A DNA sequence (GBSx1003) was identified in *S.galactiae* <SEQ ID 2873> which encodes the amino acid sequence <SEQ ID 2874>. This protein is predicted to be ABC transporter, ATP-binding protein.

- 15 Analysis of this protein sequence reveals the following:

Possible site: 56
 >>> Seems to have no N-terminal signal sequence

- 20 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2610(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

- 25 >GP:CAB49925 GB:AJ248286 ABC transporter, ATP-binding protein
 [Pyrococcus abyssi]
 Identities = 96/243 (39%), Positives = 164/243 (66%), Gaps = 2/243 (0%)
- 30 Query: 1 MIKFEHVSKVYGEKEALSDLTLSVKDGEIFGLIGHNGAGKTTTISILTSIIDATYQGQVYI 60
 MI E++ K +G KE L ++ +VKDGEI+GL+G NG+GK+TT+ IL+ II G+V +
 Sbjet: 1 MIIVENLRKRFGGKEVLKGISPTVKDGEIYGLLPNGSGKSTMRILSGIITDFEGKVIV 60
- Query: 61 DDLLETEHRDQIKKKIGYVPDSPDIFLNLTAEEYWFYFLAKIYDVAPEDIEARITKLVDIF 120
 + + + Q+K+ +GYVP++P ++ +LT E++ F+ + + + +E R+ KLV+ F
 35 Sbjet: 61 GGEVAKDPLQVKRIVGYVPETPALYESLTPAEFFSFVGGVGRGIPKDILEERVRLVEAF 120
- Query: 121 BLEEQRYNPIESFSHGMQKVIVIGALLPNPDIWILDEPLTGLDPOASFDLKEMMKHAK 180
 E+++ I + S G +QK+ +I +LL +P + ILDE + GLDP+++ +E++ E +
 40 Sbjet: 121 EIKKYNQLIGTLSFGTKQKISLISLLHDPKVLILDEAMNGLPKSARIFRELLYEFKE 180
- Query: 181 NGKTIVFSTHVLAVAEQLCDRIGILKQGKLI FVGS LGELKMKYPDKDLETIYLELAGRQA 240
 GK+++FSTHVLAAE +CDR+GI+ QG++I G++ ELK ++ LE ++L+L QA
 Sbjet: 181 EGKSIVFSTHVLALAEICDRVGIYQGRITAEGTVEELKEISKEERLEDVFLKLT--QA 238
- 45 Query: 241 SRE 243
 E
 Sbjet: 239 KEE 241

- 50 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2875> which encodes the amino acid sequence <SEQ ID 2876>. Analysis of this protein sequence reveals the following:

Possible site: 56
 >>> Seems to have no N-terminal signal sequence

- 55 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2723(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

-1036-

Identities = 182/244 (74%), Positives = 215/244 (87%)

Query: 1 MIKFEHVSQVYGEKEALSDDLTVSKDGEIFGLIGHNGAGKTTTISILTSIIDATYGVYI 60
 MI+F+HVSQ+YQ+KEALSDDL +++ DGEIFGLIGHNGAGKTTTISILTSII+A+YQ+V++
 5 Sbjct: 1 MIEFKHVSQVYGEKEALSDDLTVSKDGEIFGLIGHNGAGKTTTISILTSII+EASYGEVVFV 60

Query: 61 DDLLETHRDQIKKKIGYVPDSDIFLNLTAEEYWYFLAKIYDVAPEDIEARITKLVDIF 120
 D LLTE+R+ IKK+I YVPDSDIFLNL EYW FLAKIY V+ ED E R+ +L +F
 10 Sbjct: 61 DGQLLTENREAIKKQIAYVPDSDIFLNLTPNEYWQFLAKIYGVSDREERLAQLTTLF 120

Query: 121 ELEEQRYNPIESFSHGMQRQVIVIGALLPNPDIWILDEPLTGLDPQASFDLKEMMKHA 180
 EL+E+ I+SFSHGMQRQVIVIGAL+ NP+TWILDEPLTGLDPQASFDLKEMMK HA
 15 Sbjct: 121 ELKEEVNQITDSFSHGMQRQVIVIGALVSNPNIWILDEPLTGLDPQASFDLKEMMKAHAA 180

Query: 181 NGKTVIFSTHVLAVAEQLCDRIGILKQKGLIFVGSGLBLKMKYPDKDLETIYLELAGRQA 240
 +G TV+FSHVL+VAEQLCDRIGILK+GKLI+VGV++ ELK +PDKDLE+IYLELAGR+A
 20 Sbjct: 181 SGHTVLFSTHVLVAEQLCDRIGILKKGKLI+VGTIDELKEHHPDKDLESIYLELAGRKA 240

Query: 241 SREG 244
 EG
 20 Sbjct: 241 QEEG 244

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

25 Example 946

A DNA sequence (GBSx1004) was identified in *S.agalactiae* <SEQ ID 2877> which encodes the amino acid sequence <SEQ ID 2878>. Analysis of this protein sequence reveals the following:

Possible site: 55
 >>> Seems to have no N-terminal signal sequence

30	INTEGRAL	Likelihood = -13.43	Transmembrane	504 - 520 (495 - 529)
	INTEGRAL	Likelihood = -12.58	Transmembrane	427 - 443 (400 - 449)
	INTEGRAL	Likelihood = -10.99	Transmembrane	151 - 167 (144 - 179)
	INTEGRAL	Likelihood = -8.44	Transmembrane	194 - 210 (189 - 214)
	INTEGRAL	Likelihood = -7.96	Transmembrane	48 - 64 (46 - 68)
35	INTEGRAL	Likelihood = -7.32	Transmembrane	350 - 366 (348 - 378)
	INTEGRAL	Likelihood = -6.69	Transmembrane	475 - 491 (474 - 501)
	INTEGRAL	Likelihood = -6.00	Transmembrane	319 - 335 (318 - 337)
	INTEGRAL	Likelihood = -5.73	Transmembrane	252 - 268 (244 - 271)
	INTEGRAL	Likelihood = -4.78	Transmembrane	125 - 141 (121 - 148)
40	INTEGRAL	Likelihood = -4.51	Transmembrane	76 - 92 (71 - 98)
	INTEGRAL	Likelihood = -3.56	Transmembrane	406 - 422 (400 - 426)

----- Final Results -----

45	bacterial membrane	--- Certainty=0.6371(Affirmative) < succ>
	bacterial outside	--- Certainty=0.0000(Not Clear) < succ>
	bacterial cytoplasm	--- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2879> which encodes the amino acid sequence <SEQ ID 2880>. Analysis of this protein sequence reveals the following:

Possible site: 37
 >>> Seems to have no N-terminal signal sequence

55	INTEGRAL	Likelihood = -14.33	Transmembrane	167 - 183 (158 - 193)
	INTEGRAL	Likelihood = -12.52	Transmembrane	524 - 540 (508 - 546)
	INTEGRAL	Likelihood = -10.93	Transmembrane	63 - 79 (60 - 84)
	INTEGRAL	Likelihood = -8.39	Transmembrane	421 - 437 (414 - 456)
	INTEGRAL	Likelihood = -8.23	Transmembrane	208 - 224 (203 - 228)
	INTEGRAL	Likelihood = -8.23	Transmembrane	504 - 520 (493 - 521)
	INTEGRAL	Likelihood = -7.59	Transmembrane	139 - 155 (134 - 162)
60	INTEGRAL	Likelihood = -6.64	Transmembrane	261 - 277 (257 - 287)

-1037-

INTEGRAL Likelihood = -4.99 Transmembrane 446 - 462 (444 - 464)
 INTEGRAL Likelihood = -4.25 Transmembrane 369 - 385 (367 - 387)
 INTEGRAL Likelihood = -0.80 Transmembrane 87 - 103 (87 - 104)
 INTEGRAL Likelihood = -0.11 Transmembrane 334 - 350 (334 - 350)

5

----- Final Results -----

bacterial membrane --- Certainty=0.6731(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

10

A related sequence was also identified in GAS <SEQ ID 9173> which encodes the amino acid sequence <SEQ ID 9174>. Analysis of this protein sequence reveals the following:

Possible site: 51

>>> Seems to have no N-terminal signal sequence

15

INTEGRAL Likelihood = -14.33 Transmembrane 153 - 169 (144 - 179)
 INTEGRAL Likelihood = -12.52 Transmembrane 510 - 526 (494 - 532)
 INTEGRAL Likelihood = -10.93 Transmembrane 49 - 65 (46 - 70)
 INTEGRAL Likelihood = -8.39 Transmembrane 407 - 423 (400 - 442)
 INTEGRAL Likelihood = -8.23 Transmembrane 194 - 210 (189 - 214)
 INTEGRAL Likelihood = -8.23 Transmembrane 490 - 506 (479 - 507)
 INTEGRAL Likelihood = -7.59 Transmembrane 125 - 141 (120 - 148)
 INTEGRAL Likelihood = -6.64 Transmembrane 247 - 263 (243 - 273)
 INTEGRAL Likelihood = -4.99 Transmembrane 432 - 448 (430 - 450)
 INTEGRAL Likelihood = -4.25 Transmembrane 355 - 371 (353 - 373)
 INTEGRAL Likelihood = -0.80 Transmembrane 73 - 89 (73 - 90)
 INTEGRAL Likelihood = -0.11 Transmembrane 320 - 336 (320 - 336)

20

25

----- Final Results -----

bacterial membrane --- Certainty=0.673(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

30

An alignment of the GAS and GBS proteins is shown below.

Identities = 255/542 (47%), Positives = 378/542 (69%), Gaps = 12/542 (2%)

35

Query: 1 MNWSRIWELVKINILYSNPQTLSALRKKQEKHPKKEFSAYKSMFRNQLFQILLFSIIVF 60
 MNWS IWEL+KINILYSNPQ+L+ L+K+QEKHPK+ F AYKSM R Q I +F +IY+F
 Sbjct: 15 MNWSTIWELIKINILYSNPQSLANLKKRQEKHPKENFKAYKSMRQQALMIAMFLVIYLF 74

40

Query: 61 LEFVSLDFKEYPGYFTFYIGIFTLVSIISFIAMYSVFYESDDVKQYAYLPKSEELYVAK 120
 +F+ +DF YPG F+F + +F ++S + +F ++Y++FYES+D+K Y +LP+ SEELY+AK
 Sbjct: 75 MFIGVDFSHYPGLFSFDVAMFFIMSTLTAFSSLYTIFYESNDLKLYIHLPTSEELYIAK 134

45

Query: 121 IFATFGMSVTFLMPILTLMIYAVWRIIGGPLAVLLAIINFAILFLSVTVISLYINSLIGR 180
 I ++ GM FLMP+++L+++AYW+++G PL++L+AI+ F +L +S V+++YIN+ +G+
 Sbjct: 135 IVSSIGMGAVFLMPLISLLLIAYWQLLGNPLSILVAIVLFLVLLVSSMVLAIYINAWVGK 194

50

Query: 181 AIIRSANRKLITISLISLATFGAIVPLLFVNMSTQK--MVQGLQDIAPIPYVRGYDIDV 238
 I+RS RKLITII++ ++TFGA V + +N+++ K M G D IPY +G+YD+V
 Sbjct: 195 IIVRSRKRKLITIMFVSTFGAFVLIFAINISNNKRTMTDGVFTDYPTIPYFKGFYDVV 254

55

Query: 239 TAPFSMESLNNYYLPLLIILFLIGAIYKWMPRYQELLY---GQVKORK--VHRQIDF 292
 APFS +LLN++LPLL+IL ++ I VMP YY+E Y +VKQ K V+R
 Sbjct: 255 QAPFSTAALLNFWLPLLILAMVYGVITKVMPTYYREAFYISNENKVKQTKKPVNRP--- 311

60

Query: 293 SKRESINKTLVKHHLSSLQNTLLTNTFLMPLLYLAMFIVPIINNGKEIGRFFNENYFGI 352
 + +S+ + L KHL +LQNTLLT T+LMEL+Y+ +FI P L+ G + + +YFG+
 Sbjct: 312 HQNQSLAQLLRKHLLTLQNTLLTQTYLMPLMYVMLFIGPSLSRGTGFFKHISPDYFGV 371

65

Query: 353 AFLAGILIGSLCVMFASIVGVGISLEKSIFYFIKSLPISFSYFLKHKFVTLITLQLAVPT 412
 A L G+ +G +C P S +GVGISLEK NF FIKSLPI+ FL KF L+ LQL VP
 Sbjct: 372 ALLFGVSLGVMCATPTSFIVGVGISLEKDNFTFIKSLPITLKKFLMDKFCLLVGLQLIVPM 431

Query: 413 FIYFLVGFFLLKLSILVLLSFILGLVFMGLIEGQFIYRRDYKHLFLNWQEVTLQFNRLG 472
 IY + G F+L L L+ ++F LG +++G+ +YRRDY+ L L WQ++TQLF RG G

-1038-

Sbjct: 432 VIYLVFGLFVLHLHPLLTIAFCLGYALSILVQGLMYRRDYRLDLKWQDMTQLFTRGDG 491

Query: 473 QWLLVGSLEFGMMIIGSFL-IGISIFWSMVWNTVAVNIIILIGLLILSICQYLLLNKFWK 531

5 Sbjct: 492 QWLTMLIFGNLIVAGVLGFGAVIIANIIQQPLLISILLSCLILMVLGLAQLWIQKTFWK 551

Query: 532 KL 533

L

10 Sbjct: 552 SL 553

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 947

15 A DNA sequence (GBSx1005) was identified in *Sagalactiae* <SEQ ID 2881> which encodes the amino acid sequence <SEQ ID 2882>. Analysis of this protein sequence reveals the following:

Possible site: 44

>>> Seems to have no N-terminal signal sequence

20 INTEGRAL Likelihood = -8.12 Transmembrane 242 - 258 (239 - 265)
 INTEGRAL Likelihood = -7.64 Transmembrane 430 - 446 (421 - 450)
 INTEGRAL Likelihood = -5.84 Transmembrane 120 - 136 (113 - 139)
 INTEGRAL Likelihood = -5.52 Transmembrane 212 - 228 (210 - 232)
 INTEGRAL Likelihood = -5.20 Transmembrane 287 - 303 (283 - 313)
 INTEGRAL Likelihood = -3.56 Transmembrane 148 - 164 (143 - 166)
 25 INTEGRAL Likelihood = -0.48 Transmembrane 382 - 398 (382 - 398)

----- Final Results -----

30 bacterial membrane --- Certainty=0.4248(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB15963 GB:Z99124 phosphotransferase system (PTS)

beta-glucoside-specific enzyme IIABC component [Bacillus subtilis]

35 Identities = 175/447 (39%), Positives = 266/447 (59%), Gaps = 10/447 (2%)

Query: 4 EYITLSKNIKHLGGQNNINNVYHCQTRLRFSLNDPTKVNLEQLKTLKEVKTVVISGGQH 63

+Y LSK+I++ +GG+ N+ V HC TRLRF+L+D K + QL+ L V ISG Q

Sbjct: 2 DYDKLSKDILQLVGGEENVQRVHICMTRLRFNLHDNAKADRSQLEQLPGVMGTNISGEQF 61

40 Query: 64 QIVIGTHVAKVFEEI---NSLIETNSTTKIEQTKKAKAVSRIIDFVSGTFQPILPALSGA 120

QI+IG V KV++ I ++L + S Q K +S + D +SG F PILPA++GA

Sbjct: 62 QIIIGNDVPKVYQAIVRHSNLSDEKSAGSSSQKKNV--LSAVFDVISGVFTPIIPAIAGA 119

45 Query: 121 GMIKALLALLLVFKILTPSSQTYILNLNLFADGVFYFLPILIAITAAQKLKANPILALGTV 180

GMIK L+AL + F + SQ +++L DG FYFLP+L+A++AA+K +NP +A

Sbjct: 120 GMIKGLVALAVTFGWMAEKSQVHVILTAVGDFYFLPLLAMSAARKFGSNPYVAAAIA 179

Query: 181 VMLLHPNWNANLVASGKPVSLFHTIPFTLTNYASSVIPIILIICVQAYIEKYLKQIIPKSL 240

+LHP+ L+ +GKP+S F +P T Y+S+VIPI+L I + +Y+EK++ + SL

50 Sbjct: 180 AAILHPDLTALLGAGKPIS-FIGLPVTAATYSSTVIPILLSIWIASYVEKWIDRFTHASL 238

Query: 241 RLVLVPMFLIFLSMGLSFSILGPMGTIAGQYLAVIFTFLSKYASW-APAFIVGAFAPILI 299

+L++VP L + L+ +GP+G I G+YL+ +L +A A FL G F+ ++I

55 Sbjct: 239 KLIVVPTFTLLIVVPLTLITVGPLGAILGEYLSSGVNYLFDHAGLVAMIFLAGTFS-LII 297

Query: 300 MFGVHSGIAALGITQLAKLGVDISIFGPGMLCSNIAQATAGTVVTLITKEKKLKEIAGPAA 359

M G+H + I +A+ G D + P M +N+ QA A V L ++ KK K +A +

Sbjct: 298 MTGMHYAFVPIMINNIAQNGHDYLL-PAMFLANMGQAGASFAVFLRSRNKKFKSLALITS 356

60 Query: 360 ITAYMGITEPILYGVNLPKRYPLIASLIGGGGLGYAGIMNAHRFAV-GSSGLPGLFLYI 418

ITA MGITEP +YGVN+ + P A+LIGG GG + G+ + V G++GLP + ++I

Sbjct: 357 ITALMGITEPAMYGVNMLKKPFAALIGGAAGGAFYGMTGVASYIVGGNAGLPSIPVFI 416

-1039-

Query: 419 SHTSTHLFITMLIAVIITVSTTAILTF 445
 T + I ++IA S +L F
 Sbjct: 417 GPTFIYAMIGLVIAFAAGTSAAYLLGF 443

5

There is also homology to SEQ ID 2884.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 948

- 10 A DNA sequence (GBSx1006) was identified in *S.agalactiae* <SEQ ID 2885> which encodes the amino acid sequence <SEQ ID 2886>. This protein is predicted to be gamma-glutamyl kinase (proB). Analysis of this protein sequence reveals the following:

Possible site: 58
 >>> Seems to have no N-terminal signal sequence
 15 INTEGRAL Likelihood = -0.11 Transmembrane 160 - 176 (160 - 176)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.1044(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 20 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA63147 GB:X92418 gamma-glutamyl kinase [Streptococcus thermophilus]
 Identities = 200/265 (75%), Positives = 235/265 (88%)
 25 Query: 1 MKRHFETTRIVIKVGTSSLVQTSKINLSKIDHLAFVISSLMNRGMEVILVSSGAMGFG 60
 MKR+P++ +R+VIK+GTSSLV SCKINL KID LAFVISSL N+G+EV+LVSSGAMGFG
 Sbjct: 1 MKRNFDSVKRLVIKIGTSSLVLP SGKINLEKIDQLAFVISSLHNKGIEVVLVSSGAMGFG 60
 30 Query: 61 LDILKMDKRPQEISQQQAVSSVGQVAMMSLYSQIFSHYQTHVSQILLTRDVVVPESLQN 120
 L++L ++KRP E+ +QQAVSSVGQVAMMSLYSQ+FSHYQT VSQ+LLTRDVV + ESL N
 Sbjct: 61 LNVLDLEKRPAEVKGQQAVSSVGQVAMMSLYSQVFSHYQTKVSQLLLRDVVEYSESLAN 120
 35 Query: 121 VTNSFESLLSMGILPIVNENDAVSVDEMHDHKTFGDNDRLSAVAVAKITKADLLIMLSDDID 180
 N+FESL +G++PIVNENDAVSVDEMHDH TKFGDNDRLSA+VAK+ ADLLIMLSDDID
 Sbjct: 121 AINAFESLFELGVVPIVNENDAVSVDEMHDHATKFGDNDRLSAIVAKVVGADLLIMLSDDID 180
 40 Query: 181 GLFDKPNPNYDDAVLRSHVSEITDDIIKSAGGAGSKFGTGGMLSKIKSAQMVFDNNGQMI 240
 GLFDKPNPN+Y+DA LRS+V EIT++I+ SAGGAGSKFGTGGM+SKIKSAQMVFN QM+
 Sbjct: 181 GLFDKPNPNVYEDATLRSYVPEITEIILASAGGAGSKFGTGGMMSKIKSAQMVFNQSQMV 240
 45 Query: 241 LMNGANPRDILKVLGDGHNIGTYFAQ 265
 LMNG NPRDIL+VL+G IGT F Q
 Sbjct: 241 LMNGENPRDILRVLEGAKIGTLFKQ 265

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2887> which encodes the amino acid sequence <SEQ ID 2888>. Analysis of this protein sequence reveals the following:

Possible site: 61
 >>> Seems to have no N-terminal signal sequence
 50 INTEGRAL Likelihood = -1.97 Transmembrane 163 - 179 (163 - 179)
 INTEGRAL Likelihood = -0.06 Transmembrane 124 - 140 (124 - 140)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.1786(Affirmative) < succ>
 55 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

-1040-

>GP:CAA63147 GB:X92418 gamma-glutamyl kinase [Streptococcus thermophilus]
Identities = 212/265 (80%), Positives = 237/265 (89%)

5 Query: 4 MKRQFEDVTRIVIKIGTSSLVLTGKINLEKIDQLAFVISSLMNKGKEVILVSSGAMGFG 63
MKR F+ V R+VIKIGTSSLVLP+GKINLEKIDQLAFVISSL NKG EV+LVSSGAMGFG
Sbjct: 1 MKRNFDSVKRLVIKIGTSSLVLP+GKINLEKIDQLAFVISSLMNKGKEVILVSSGAMGFG 60

10 Query: 64 LDILKMEKRPTNLAQQAVSSVGQVAMMSLYSQIFAYYQTNVSQILLTRDVVVPFESLAN 123
L++L +EKRP + KQAVSSVGQVAMMSLYSQ+F++YQT VSQ+LLTRDVV + ESLAN
Sbjct: 61 LNVLDLEKRPAAEVGKQAVSSVGQVAMMSLYSQVFSHYQTKVSQILLTRDVVVESESAN 120

15 Query: 124 VTNAFESLISLGIPIVNENDAVSVDEMHDATKFGDNDRLSAVVAGITKADLLIMLSID 183
NAFESL LG+VPIVNENDAVSVDEMHDATKFGDNDRLSA+VA + ADLLIMLSID
Sbjct: 121 AINAFESLFELGVVPIVNENDAVSVDEMHDATKFGDNDRLSAIVAKVVGADLLIMLSID 180

20 Query: 184 GLFDKNPTIYEDALRSHVANITQEIISAGGAGSKFGTGGMLSKVQSAQMVFENKGMV 243
GLFDKNP +YEDA LRS+V IT+EI+ASAGGAGSKFGTGGM+SK++SAQMVFEN+ QMV
Sbjct: 181 GLFDKNPNVYEDATLRSYVPEITEILASAGGAGSKFGTGGMMSKIKSAQMVFENQSQMV 240

25 Query: 244 LMNGANPRDILRVLEGQPLGTWFKQ 268
LMNG NPRDILRVLEG +GT FKQ
Sbjct: 241 LMNGENPRDILRVLEGAKIGTLFKQ 265

An alignment of the GAS and GBS proteins is shown below.

25 Identities = 217/265 (81%), Positives = 242/265 (90%)

30 Query: 1 MKRHFETTRIVIKVGTSSLVQTSKINLSKIDHLAFVISSLMNRGMEVILVSSGAMGFG 60
MKR FE RIVIK+GTSSLV +GKINL KID LAFVISSLMN+G EVILVSSGAMGFG
Sbjct: 4 MKRQFEDVTRIVIKIGTSSLVLTGKINLEKIDQLAFVISSLMNKGKEVILVSSGAMGFG 63

35 Query: 61 LDILKMDKRPQEISQQAVSSVGQVAMMSLYSQIFSHYQTHVSQILLTRDVVVPFESLQN 120
LDILKM+KRP +++QAVSSVGQVAMMSLYSQIF++YQT+VSQILLTRDVVVPFESL N
Sbjct: 64 LDILKMEKRPTNLAQQAVSSVGQVAMMSLYSQIFAYYQTNVSQILLTRDVVVPFESLAN 123

40 Query: 121 VTNSFESLLSMGILPIVNENDAVSVDEMHDHKTGFGDNDRLSAVVAKITKADLLIMLSID 180
VTN+FESL+S+GI+PIVNENDAVSVDEMDH TKFGDNDRLSAVVA ITKADLLIMLSID
Sbjct: 124 VTNAFESLISLGIPIVNENDAVSVDEMHDATKFGDNDRLSAVVAGITKADLLIMLSID 183

45 Query: 181 GLFDKNPNYDDAVLRSHVSEITDDIISAGGAGSKFGTGGMLSKIKSAQMVFENKGMV 240
GLFDKNP IY+DA LRSHV+ IT +II SAGGAGSKFGTGGMLSK++SAQMVF+N QGM+
Sbjct: 184 GLFDKNPTIYEDALRSHVANITQEIISAGGAGSKFGTGGMLSKVQSAQMVFENKGMV 243

Query: 241 LMNGANPRDILKVLGDHNGITYFAQ 265
LMNGANPRDIL+VL+G +GT+F Q
Sbjct: 244 LMNGANPRDILRVLEGQPLGTWFKQ 268

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 949

50 A DNA sequence (GBSx1007) was identified in *S.agalactiae* <SEQ ID 2889> which encodes the amino acid sequence <SEQ ID 2890>. This protein is predicted to be unnamed protein product (proA). Analysis of this protein sequence reveals the following:

Possible site: 55

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3517(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2891> which encodes the amino acid sequence <SEQ ID 2892>. Analysis of this protein sequence reveals the following:

Possible site: 55
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:CAA63148 GB:X92418 gamma-glutamyl phosphate reductase
 [Streptococcus thermophilus]
 Identities = 309/416 (74%), Positives = 355/416 (85%)
 Query: 1 MTDMRRLGQRAKQASLLIAPLSTQIKNRFSLTAKALVDDTQTLLAANQKDLANAKEHGI 60
 MT + LGQ+AK AS IA LST KN L+ +AKALV ++ + N KD+ANA E+GI
 Sbjct: 1 MTYVDTLGQAKVASRQIAKLSTAAKNDLLNQAKALVAESDYIFTENAKDMANASENGI 60
 Query: 61 SDIMMDRLRLTSERIKAIAGVQVQVADLADFIGQVIKGYTNLDGLKILQKRVPLGVAMI 120
 S IM DRL LT +RI IA+GV+QVADL DFIGQV++GYTNLDGLKI+QKRVPL+GVAMI
 Sbjct: 61 SKIMQDRLRLLTEDRIAGIAGVGRQVADLQDFIGQVVRGYTNLDGLKIVQKRVPMGVAMI 120
 Query: 121 FESRPNVSVDAFSLAFKTNNAIILRGGKDALHSNKALVKLIRQSLEKSGITPDAVQLVED 180
 FESRPNVS+DAFSLAFKTNNAIILRGG+DA++SNKALV + R++L+ +GIT DAVQ VED
 Sbjct: 121 FESRPNVSIDAFSLAFKTNNAIILRGGRDAINS NKALVTVARKALKNAGITADAVQFVED 180
 Query: 181 PSHAVAEELMQATDYVDVLIIPRGGAQLIQTVKEKAKVPVIETGVGNVHIYVDAQADLDIA 240
 SH VAEELM AT YVD+LIPRGA+LIQTVKEKAKVPVIETGVGN HIYVD A+LD+A
 Sbjct: 181 TSHEVAEELMVATKYVDLLIIPRGGARLIQTVKEKAKVPVIETGVGNCHIYVDKYANLDMA 240
 Query: 241 TKIVINAKTKRPSVCNAAEGLVIHEAARFIPMLEKAINQVQVPEWRADDKALPLFEQA 300
 T+IVINAKT+RPSVCNAAE LV+H + F+P LEKAI+++Q VE+RAD++AL L E+A
 Sbjct: 241 TQIVINAKTQRPSVCNAAESLVVHADIVEEFLPNLEKAISKIQSVEFRADERALKLMEKA 300
 Query: 301 VPAKAEDFETEFLDYIMSVKVVSSLEEAIWINQYTSHHSEAIITRDIKAAETFDQLVDA 360
 VPA EDF TEFLDYIMSVKVV SL+EAI+WIN YT+ HSEAI+T+DI AE FQD VDA
 Sbjct: 301 VPASPEDFATEFLDYIMSVKVVDSLEAINWINTYTTSHSEAIIVTQDISRAEQFQDDVDA 360
 Query: 361 AAVYVNASTRFTDGFVFLGAEIGISTQKMHARGPMGLEALTSTKIFYINGDGHIRE 416
 AAVYVNASTRFTDGFVFLGAEIGISTQKMHARGPMGLEALTSTKIFYING G IRE
 Sbjct: 361 AAVYVNASTRFTDGFVFLGAEIGISTQKMHARGPMGLEALTSTKIFYINGQGQIRE 416

An alignment of the GAS and GBS proteins is shown below.

Identities = 307/417 (73%), Positives = 353/417 (84%), Gaps = 1/417 (0%)
 Query: 1 MTYIEILGQNAKKASQSVARLSTASKNEILRDLARNIVADTETILTENARDVVKADNGI 60
 MT + LGQ AK+AS +A LST KN L LA+ +V DT+T+L N +D+ AK++GI
 Sbjct: 1 MTDMRRLGQRAKQASLLIAPLSTQIKNRFSLTAKALVDDTQTLLAANQKDLANAKEHGI 60
 Query: 61 SEIMVDRRLRNKDRIQAIANGIYQVADLADFIGQVVSQYTNLDGLKILKKRVPLGVAMI 120
 S+IM+DRLRL +RI+AIA G+ QVADLADFIGQV+ GYTNLDGLKIL+KRVPLGVAMI
 Sbjct: 61 SDIMMDRLRLTSERIKAIAGVQVQVADLADFIGQVIKGYTNLDGLKILQKRVPLGVAMI 120
 Query: 121 FESRPNVSVDAFSLAFKTNNAIILRGGKDAIFSNALVNCMRQTLQDTGHNPDIVQLVED 180
 FESRPNVSVDAFSLAFKT NAIILRGGKDA+ SN ALV +RQ+L+ +G PD VQLVED
 Sbjct: 121 FESRPNVSVDAFSLAFKTNNAIILRGGKDALHSNKALVKLIRQSLEKSGITPDAVQLVED 180
 Query: 181 TSHVVAEELMQATDYVDVLIIPRGGAQLIQTVKEKSKIPVIETGVGNVHIYIDEFADLDMA 240
 SH VAEELMQATDYVDVLIIPRGGAQLIQTVKEK+K+PVIETGVGNVHIY+D ADLD+A
 Sbjct: 181 PSHAVAEELMQATDYVDVLIIPRGGAQLIQTVKEKAKVPVIETGVGNVHIYVDAQADLDIA 240
 Query: 241 AKIVINAKTQRPSVCNAAEGLVHVQAIAGFLSQLEKMLKESNQSVFEFRADDEALQLEN 300
 KIVINAKT+RPSVCNAAEGLV+H+A+A F+ LEK + + Q VE+RAD++AL L E

-1042-

Sbjct: 241 TKIVINAKTKRPSVCNAAEGLVIHEAVAARFIPMLEKAINQV-QPVEWRADDKALPLFEQ 299

Query: 301 AVAASESDYATEFLDYIMSVKVVDSFEQAISWINKYSSHHSEAITTNISRAEIFQDMVD 360
AV A D+ TEFLDYIMSVKVV S E+AISWIN+Y+SHHSEAIT +I AE FQD+VD

Sbjct: 300 AVPAKAEDFETFLDYIMSVKVVSSLEFAISWINQYTSHHSEAITTRDIKAAETFDQLVD 359

Query: 361 AAAYVYNASTRFTDGFVFGGLGAEIGISTQKLHARGPMGLEALTSTKYINGTGQVRE 417
AAAYVYNASTRFTDGFVFGGLGAEIGISTQK+HARGPMGLEALTSTK+YING G +RE

Sbjct: 360 AAAYVYNASTRFTDGFVFGGLGAEIGISTQKMHARGPMGLEALTSTKFYINGDGHIRE 416

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 950

A DNA sequence (GBSx1008) was identified in *Sagalactiae* <SEQ ID 2893> which encodes the amino acid sequence <SEQ ID 2894>. Analysis of this protein sequence reveals the following:

Possible site: 53

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1859(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9531> which encodes amino acid sequence <SEQ ID 9532> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2895> which encodes the amino acid sequence <SEQ ID 2896>. Analysis of this protein sequence reveals the following:

Possible site: 23

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0853(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 259/315 (82%), Positives = 287/315 (90%)

Query: 1 MTNDFHHITVLLHETVDMLDIKPDGIYVDATLGGAGHSEYLLSQLGPDGHLYAFDQDQKA 60

MT +FHH+TVLLHETVDMLDIKPDGIYVDATLGG+GHS YLLS+LG +GHLY FDQDQKA

Sbjct: 22 MTKEFHHVTVLLHETVDMLDIKPDGIYVDATLGGSGHSAYLLSKLGEEGHLYCFDQDQKA 81

Query: 61 IDNAHIRLKKYVDTGQVTFIKDNFRNLSSNLKALGVSEINGICYDLGVSSPQLDERERGF 120

IDNA + LK Y+D GQVTFIKDNFR+L + L ALGV EI+GI YDLGVSSPQLDERERGF

Sbjct: 82 IDNAQVTLKSYIDKGQVTFIKDNFRHLKARLTALGVDEIDGILYDLGVSSPQLDERERGF 141

Query: 121 SYKQDAPLDMRMNREQLTAYDVVNTYSYHDLVRIFFKYGEDKFSKQIARKIEQVRAEKT 180

SYKQDAPLDMRM+R+ LTAY+VVNTY ++DLV+IFFKYGEDKFSKQIARKIEQ RA K

Sbjct: 142 SYKQDAPLDMRMDRQSLLTAYEVVNTYPFNDLVKIFFKYGEDKFSKQIARKIEQARAIPK 201

Query: 181 ISTTTTELAEIIKSSSAKELKKKGHPAKQIFQAIRIEVNDELGAADESIQAMDLLAVDG 240

I TTTTELAE+IK++K AKELKKKGHPAKQIFQAIRIEVNDELGAADESIQ AM+LLA+DG

Sbjct: 202 IETTTTELAEIIKAAKPAKELKKKGHPAKQIFQAIRIEVNDELGAADESIQDAMELLALDG 261

Query: 241 RISVITFHSLEDRLTKQLFKEASTVEVPKGLPFIPDDLQPKMELVNRKPILPSQEELEAN 300

RISVITFHSLEDRLTKQLFKEASTV+VPKGLP IP+D++PK ELV+RKPILPS EL AN

Sbjct: 262 RISVITFHSLEDRLTKQLFKEASTVDVPKGLPLIPEDMKPKFELVSRKPILPSHSELTAN 321

-1043-

Query: 301 NRAHSAKLRVARRIR 315
RAHSAKLRVA++IR
Sbjct: 322 KRAHSAKLRVAKKIR 336

- 5 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 951

- 10 A DNA sequence (GBSx1009) was identified in *S.agalactiae* <SEQ ID 2897> which encodes the amino acid sequence <SEQ ID 2898>. This protein is predicted to be FtsL. Analysis of this protein sequence reveals the following:

Possible site: 42
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -8.92 Transmembrane 30 - 46 (24 - 49)

- 15 ----- Final Results -----
bacterial membrane --- Certainty=0.4567(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

- 20 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC95455 GB:AF068903 Y11D [Streptococcus pneumoniae]
Identities = 44/99 (44%), Positives = 71/99 (71%)

- 25 Query: 5 KRTEAVTQTLQRHIKTFSRIEKAFYGAIVITAIIMAVGIIYLSNSLQVKQEVNQLNSKI 64
++ E Q LQ +K FSR+EKAFY +I +T +I+A+ II++Q+ LQV+ ++ ++N++I
Sbjct: 3 EKMEKTGQILQMLKRFSRVEKAFYFSIAVTTLIVAISIIFMQTKLLQVQNDLTKINAQI 62

- Query: 65 NDKQTEFDNAKQEVNELSNRDRITKIAKDAGLTIONDNI 103
+K+TE D+AKQEVNEL +R+ +IA L + N+NI
30 Sbjct: 63 EEKTELDDAKQEVNELLRERLKEIANSHDLQLNNENI 101

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2899> which encodes the amino acid sequence <SEQ ID 2900>. Analysis of this protein sequence reveals the following:

- 35 Possible site: 50
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -5.79 Transmembrane 40 - 56 (37 - 58)
----- Final Results -----
bacterial membrane --- Certainty=0.3314(Affirmative) < succ>
40 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

- 45 >GP:AAC95455 GB:AF068903 Y11D [Streptococcus pneumoniae]
Identities = 45/94 (47%), Positives = 69/94 (72%)
Query: 24 LQKRIKTFSRIEKAFYTAIIVTAIMAVSIIYLSRKLQLQEQEITSLNSHISDQKLELNN 83
LQ ++K FSR+EKAFY +I VT + +A+SII++Q++ LQ+Q ++T +N+ I ++K EL++
Sbjct: 12 LQMLKRFSRVEKAFYFSIAVTTLIVAISIIFMQTKLLQVQNDLTKINAQIEKKTELD 71
50 Query: 84 AKQEVNELSRDRIIDAGKAGLSNRNNNIKKVE 117
AKQEVNEL R +R+ +IA L N NI+ E
Sbjct: 72 AKQEVNELLRERLKEIANSHDLQLNNENIRIAE 105

- 55 An alignment of the GAS and GBS proteins is shown below.

Identities = 71/108 (65%), Positives = 87/108 (79%), Gaps = 1/108 (0%)

-1044-

Query: 1 MTNEKRTQAVTQTLQRHIKTFSEKAFYGAIVITAIIMAVGIIYLSNSLQVKQEVNQL 60
 MTNEKRT+ VT LQ+ IKTFSEKAFY AI++TAI MAV IIVLQS LQ++QE+ L
 Sbjct: 11 MTNEKRTQVVTNALQKRIKTFSEKAFYTAIIVTAITMAVSIIYLSRKLQLQEQEITSL 70

5 Query: 61 NSKINDKQTEFDNAKQEVNELSNRDRITKIAKDAGLTIQNDNIYRKVD 108
 NS I+D++ E +NAKQEVNELS RDRI IA AGL+ +N+NI +KV+
 Sbjct: 71 NSHISDQKLELNNAKQEVNELSRDRIDIAGKAGLSNRNNNI-KKVE 117

SEQ ID 2898 (GBS82) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell
 10 extract is shown in Figure 15 (lane 2; 2 bands).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 vaccines or diagnostics.

Example 952

A DNA sequence (GBSx1010) was identified in *S.agalactiae* <SEQ ID 2901> which encodes the amino
 15 acid sequence <SEQ ID 2902>. Analysis of this protein sequence reveals the following:

Possible site: 21
 >>> Seems to have no N-terminal signal sequence

20 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1435(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

25 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for
 vaccines or diagnostics.

Example 953

A DNA sequence (GBSx1011) was identified in *S.agalactiae* <SEQ ID 2903> which encodes the amino
 30 acid sequence <SEQ ID 2904>. This protein is predicted to be unnamed protein product. Analysis of this
 protein sequence reveals the following:

Possible site: 47
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood =-13.90 Transmembrane 37 - 53 (30 - 60)

35 ----- Final Results -----
 bacterial membrane --- Certainty=0.6562(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

40 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2905> which encodes the amino acid
 sequence <SEQ ID 2906>. Analysis of this protein sequence reveals the following:

Possible site: 42

45 >>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood =-13.06 Transmembrane 33 - 49 (24 - 53)

50 ----- Final Results -----
 bacterial membrane --- Certainty=0.6222(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 480/753 (63%), Positives = 603/753 (79%), Gaps = 8/753 (1%)

```

5  Query: 5  KKLKKIFLDYVIHIDRRSPQKNRERVGNLMILTIFFIFIINFVIVGTDSKFGVNL 64
   KK +K LDYV+ RDRR+P +NR RVGQN+M+LTIF+FFIFIFIINF+II+GTD KFGV+L
   Sbjct: 2  KKWQKYVLDYVV--RDRRTPVENRVRVGNMMLLTIFIFFIFIINFMIIGTDQKFGVSL 59

10 Query: 65  SKEAKKVYQQSMTVQAKRGTIYDRNGNPIAEDATTYSLYAIISKNYTTATGQKLYVQPSQ 124
   S+ AKKVYQ+++T+QAKRGTIYDRNG IA D+TTYS+YAI+ K++ +A+ +KLYVQPSQ
   Sbjct: 60  SEGAKKVYQETVTIQAIRGTIYDRNGTAIAVDSTTYSIYAILDKSFVSASDEKLYVQPSQ 119

15 Query: 125  YEKVASILENKLGMKKNLVLKQLNQKLFQVSFGSSGSGLSYTKMADIKKTEKSDIKGI 184
   YE VA IL+ LGMKK V+KQL +K LQVVSFG SSG+SY+ M+ I+K ME + IKGI
   Sbjct: 120  YETVADILKKHLGMKKTVDIKQLKRRKGLFQVSFGPSGSGISYSTMTIQKAMEDAKIKGI 179

20 Query: 185  GFSTSEGRIPYNGIFASQFIGF-TLPQDDGDG-KKLVGNTGLEAALNKVLSGTDGKVITYE 242
   F+TSPGR+YPNG FAS+FIG +L +D G K LVG TGLEA+ +K+LSG DG +TY+
   Sbjct: 180  AFTTSPGRMYPNGTFASEFIGLASLTEDKKTGVKSLVGKTGLEASFDKILSGQDGVITYQ 239

25 Query: 243  KDRSGNVLLGTATTERRAVNGKDIYTTLSEPIQTVLETQMDVFAEKTGKFASATVUNA 302
   KDR+G LLGT T ++A++GKDIYTTLSEPIQT LETQMDVF K+ G+ ASAT+VNAK
   Sbjct: 240  KDRNGTLLGTGKTGKKAIDGKDIYTTLSEPIQTFLETQMDVFQAKSNGQLASATLVNAK 299

30 Query: 303  TGEILATSQRPTYNPTSLKGYDKKNLGTYNLLYDNFFEPGSTMKVMTLASAIDSKHFNS 362
   TGEILAT+QRPTYN TLKG + N Y+ L N FEPGSTMKVMTLA+AIK K FN
   Sbjct: 300  TGEILATQRPTYNADTLKGLNTNYKWYSALHQQN-FEPGSTMKVMTLAAAIKDKVFNP 358

35 Query: 422  YLSKFKFGYPTFRGMLHESGGLFPSDNEVTIAMSSFGQIGVTVQVQLRAFTSISNDGVM 481
   YL+KF+FG+PTRFG+ E G+FPSDN VT AMS+FGQGI VTQ+QMLRAFT+ISN+G M
   Sbjct: 419  YLTFRFGFPTRFGLKDEDAIGFPSDNIVTQAMSAFGQGISVTQIQMLRAFTAISNNGEM 478

40 Query: 482  LQPQFISSIYDPNTGTSRTARKEVVGKPVSKAASKTRDYMVTGTDPPYGTLYA-AGAP 540
   L+PQFIS IYDPNT + RTA KE+VGKPVSK+AAS+TR YM+ VGTDP +GTLY+ P
   Sbjct: 479  LEPQFISQIYDPNTASRTANKEIVGKPVSKAASETRQYIMIGVGTDPBFGTLYSKTFGP 538

45 Query: 541  VIOVGNQSVAVKSGTAQIAQEGGGYLQ-GKNDTINSVVMVPSENEPDFIMYVTIQQPEK 599
   +I+VG+ VAVKSGTAQI E G GY G + + SVVAMVP++ PDF+MYVT+ +P+
   Sbjct: 539  IIKVGDLPVAVKSGTAQIGSEDGSGYQDGLLTINVVSVVAMVPADKPDFLIMYVTMTKPQH 598

50 Query: 600  FSITFWKDVVNPVLEQATAMKETILKPLNDSEHQTKYKLSKIVGENPGHVAEELRRNLV 659
   F FW+DVVNPVLE+A M++T+ KP ++D+ QT YKL VG+NPG + ELRRNLV
   Sbjct: 599  FGPLEWQDVVNPVLEEAYLMQDTLTKPVVSDANRQTTYKLPNFVGKNPGETSSELRRNLV 658

55 Query: 660  QPIILNGSKVSKVRPGANLAENEQLLVLTINKLTLPDYGWGSKANVEQFAKWTGIKV 719
   QP++LG GSK+ KVS +PG L EN+Q+L+L+++ E+PDYGW+K+NV+ FAKWTGI +
   Sbjct: 659  QPVVLGTGSKIKKVSHQPGQTLTENQQVLILSDRFVEVPDYGWTKSNVKTFAKWTGIDI 718

   Query: 720  TYKGSTSGKVRKQSIDVGSINKIKIKITIGD 752
   ++KG+ SG+V KQS+DVGKS+ KIKK+ IT+GD
   Sbjct: 719  SFKGTDSGRVMKQSVQVGVGSLKKIKKMTITLGD 751

```

A related GBS gene <SEQ ID 8691> and protein <SEQ ID 8692> were also identified. Analysis of this protein sequence reveals the following:

```

60  Lipop: Possible site: -1  Crend: 8
   McG: Discrim Score: -4.31
   GvH: Signal Score (-7.5): -7.07
   Possible site: 47
   >>> Seems to have no N-terminal signal sequence
65  ALOM program count: 1 value: -13.90 threshold: 0.0

```

-1046-

INTEGRAL Likelihood = -13.90 Transmembrane 37 - 53 (30 - 60)
PERIPHERAL Likelihood = 5.30 450
modified ALOM score: 3.28

5 *** Reasoning Step: 3

----- Final Results -----

10 bacterial membrane --- Certainty=0.6562(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

15 ORF00411(301 - 2556 of 2856)
GP|6779111|emb|CAB70457.1||A94911(1 - 752 of 752) unnamed protein product {unidentified},
homology to penicillin-binding protein 2x (S. pneumoniae)
%Match = 77.4
%Identity = 99.7 %Similarity = 99.9
Matches = 750 Mismatches = 1 Conservative Sub.s = 1

20 66 96 126 156 186 216 246 276
RIEKAFYGAIVITAIIMAVGIIYLSQNSLQVQKEVNQLNSKINDKQTEFDNAKQEVNELSNRDRITTKIAKDAGLTIQNND

25 306 336 366 396 426 456 486 516
IYRKVD*SVTFKKLKKIFLDYVIHIDRRSPQKNRERVGNLMILTIFFIFIINFVIVGTDSKFGVNLSEAKKVY
|||||
VTFKKLKKIFLDYVIHIDRRSPQKNRERVGNLMILTIFFIFIINFVIVGTDSKFGVNLSEAKKVY
10 20 30 40 50 60 70

30 546 576 606 636 666 696 726 756
QQSMTVQAKRGTYIDRNGNPIAEDATTYSLYAIISKNYTTATGQKLYVQPSQYKVASILENKLGMKKNLVLKQLNQKKL
|||||
QQSMTVQAKRGTYIDRNGNPIAEDATTYSLYAIISKNYTTATGQKLYVQPSQYKVASILENKLGMKKNLVLKQLNQKKL
90 100 110 120 130 140 150

35 786 816 846 876 906 936 966 996
FQVSFGSSGSGLSYTKMADIKKTMESKDIKGIGFSTSPGRIYPNGIFASQFIGFTLPQDDGDGKKLVGNTGLEAALNKVL
|||||
FQVSFGSSGSGLSYTKMADIKKTMESKDIKGIGFSTSPGRIYPNGIFASQFIGFTLPQDDGDGKKLVGNTGLEAALNKVL
170 180 190 200 210 220 230

40 1026 1056 1086 1116 1146 1176 1206 1236
SGTDGKVITYEKDRSGNVLLGTATTERRAVNGKDIYTTLSEPIQTVLETQMDVFAEKTGKGFASATVVNAKTGEILATSQR
|||||
SGTDGKVITYEKDRSGNVLLGTATTERRAVNGKDIYTTLSEPIQTVLETQMDVFAEKTGKGFASATVVNAKTGEILATSQR
250 260 270 280 290 300 310

45 1266 1296 1326 1356 1386 1416 1446 1476
PTYNPSSTLKGYDKKNLGTYNLLYDNFFEPGSTMKVMTLASAIDSKHFNSTEVYNSAQYKIADAIIRDWDVNEGLSSGSY
|||||
PTYNPSSTLKGYDKKNLGTYNLLYDNFFEPGSTMKVMTLASAIDSKHFNSTEVYNSAQYKIADAVIRDWDVNEGLSSGSY
330 340 350 360 370 380 390

50 1506 1536 1566 1596 1626 1656 1686 1716
MIFPQGFAHSSNVGMVTLQKMGDRDKWNLNLSKFKGYPTFRFGLHESGGLFPSDNEVTIAMSSFGQIGVTVQVOMLRAF
|||||
MIFPQGFAHSSNVGMVTLQKMGDRDKWNLNLSKFKGYPTFRFGLHESGGLFPSDNEVTIAMSSFGQIGVTVQVOMLRAF
410 420 430 440 450 460 470

55 1746 1776 1806 1836 1866 1896 1926 1956
TSISNDGVMLQPQFISSIIDPNTGTSRTARKEVVGKPVSKAASKTRDYMVTVGTDPPYGTLYAAGAPVIQVGNQSVAVK
|||||
TSISNDGVMLQPQFISSIIDPNTGTSRTARKEVVGKPVSKAASKTRDYMVTVGTDPPYGTLYAAGAPVIQVGNQSVAVK
490 500 510 520 530 540 550

60 1986 2016 2046 2076 2106 2136 2166 2196
SGTAQIAQEGGGYLQKNDPTINSVAMVPSENPDFIMYVTIQOPEKFSITFWKDVVNPVLEQATAMKETILKPEGLNDSE
|||||

-1047-

SGTAQIAQEGGGGYLQGNKNTINSVVAMVPSNPDFIMYVTIQPEKFSITFWKDVVNPVLEQATAMKETILKPVLNDS
570 580 590 600 610 620 630

2226 2256 2286 2316 2346 2376 2406 2436
HQT KYKL SKIV GENP GHV AEEL RRNL VQPI ILGN GSKV SKVSK R P GAN LAENE QLLV LTNK L TELP DMYG WSKAN VEQFA
|||||
HQT KYKL SKIV GENP GHV AEEL RRNL VQPI ILGN GSKV SKVSK R P GAN LAENE QLLV LTNK L TELP DMYG WSKAN VEQFA
650 660 670 680 690 700 710

2466 2496 2526 2556 2586 2616 2646 2676
KWTG IKVT YKGST SGKVR KQS IDVG KSINK IKKIK ITIGD*HVP KYNGR CHS IC PDS YCHS ALH*VLP I EEN WRATNA*R
|||||
KWTG IKVT YKGST SGKVR KQS IDVG KSINK IKKIK ITIGD
730 740 750

15 SEQ ID 8692 (GBS352d) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 145 (lane 15 & 16; MW 105.5kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 145 (lane 17 & 18; MW 80.5kDa), in Figure 182 (lane 3; MW 80kDa) and in Figure 185 (lane 4; MW 105kDa). Purified GBS352d-GST is shown in lane 5 of Figure 236.

20 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 954

A DNA sequence (GBSx1012) was identified in *S.agalactiae* <SEQ ID 2907> which encodes the amino acid sequence <SEQ ID 2908>. Analysis of this protein sequence reveals the following:

```

25      Possible site: 36
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.1950(Affirmative) < succ>
30      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*:

35 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 955

A DNA sequence (GBSx1013) was identified in *S.agalactiae* <SEQ ID 2909> which encodes the amino acid sequence <SEQ ID 2910>. This protein is predicted to be unnamed protein product (mraY). Analysis of this protein sequence reveals the following:

```

Possible site: 18
>>> Seems to have a cleavable N-term signal seq.
    INTEGRAL      Likelihood =-15.12      Transmembrane    56 - 72 ( 47 - 76)
    INTEGRAL      Likelihood =-14.70      Transmembrane    203 - 219 ( 198 - 223)
45  INTEGRAL      Likelihood = -6.69      Transmembrane    318 - 334 ( 315 - 335)
    INTEGRAL      Likelihood = -6.64      Transmembrane    83 - 99 ( 79 - 103)
    INTEGRAL      Likelihood = -5.52      Transmembrane    179 - 195 ( 175 - 197)
    INTEGRAL      Likelihood = -5.31      Transmembrane    232 - 248 ( 230 - 249)
    INTEGRAL      Likelihood = -3.08      Transmembrane    119 - 135 ( 119 - 137)
50  INTEGRAL      Likelihood = -2.87      Transmembrane    151 - 167 ( 147 - 167)
    INTEGRAL      Likelihood = -2.34      Transmembrane    254 - 270 ( 254 - 270)

```

-1048-

----- Final Results -----

5 bacterial membrane --- Certainty=0.7050(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2911> which encodes the amino acid sequence <SEQ ID 2912>. Analysis of this protein sequence reveals the following:

Possible site: 36

10 >>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -9.55	Transmembrane	52 - 68 (48 - 75)
INTEGRAL	Likelihood = -9.39	Transmembrane	175 - 191 (171 - 194)
INTEGRAL	Likelihood = -8.12	Transmembrane	30 - 46 (23 - 48)
INTEGRAL	Likelihood = -6.37	Transmembrane	121 - 137 (119 - 145)
15 INTEGRAL	Likelihood = -6.32	Transmembrane	293 - 309 (287 - 309)
INTEGRAL	Likelihood = -5.31	Transmembrane	204 - 220 (202 - 221)
INTEGRAL	Likelihood = -5.20	Transmembrane	151 - 167 (150 - 170)
INTEGRAL	Likelihood = -4.67	Transmembrane	226 - 242 (224 - 244)
20 INTEGRAL	Likelihood = -0.11	Transmembrane	91 - 107 (91 - 107)

----- Final Results -----

25 bacterial membrane --- Certainty=0.4821(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:CAB70458 GB:A94911 unnamed protein product [unidentified]

Identities = 244/309 (78%), Positives = 273/309 (87%), Gaps = 1/309 (0%)

30 Query: 1 LKKIGGQQMHEDVKQHLAKAGTPTMGGTVFLVAVSLVSLF-SIKNTQSLALISGIL 59
 LKKIGGQQMHEDVKQHLAKAGTPTMGGTVFL+VA VSL+ S+ S+N+ +L GIL
 Sbjct: 28 LKKIGGQQMHEDVKQHLAKAGTPTMGGTVFLVALLVSLIFSIIILSKENSGNLGATFGIL 87

35 Query: 60 SIVVIYGIIGFLDDFLKIFKQINEGLTAKQKLALQLVGGLMFYFLHVSPSGISSINVFGY 119
 S+V+YGIIGFLDDFLKIFKQINEGLT KQK++LQL+ GL+FYF+HV PSG S+IN+FG+
 Sbjct: 88 SVVLIYGIIGFLDDFLKIFKQINEGLTPKQKMSLQLIAGLIFYFVHVLPSTGSAINIFGF 147

40 Query: 120 QLPLGIFYLFFVLFVWVGFSNAVNLTGIDGLASISVVISLVTYGVIAVQSQFDVLLLI 179
 L +G Y FVLFVWVGFSNAVNLTGIDGLASISVVISL+TYG+IAY Q+QFD+LL+I
 Sbjct: 148 NLEVGYLYAFFVLFVWVGFSNAVNLTGIDGLASISVVISLITYGIIAYNQTFDILLII 207

45 Query: 180 GAMIGALLGFFCFNHKPAKVFMGDVGSALGAMLAASIALRQEWILLIIGIVVLETSS 239
 MIGALLGFF FNHKAQVFMGDVGSALGAMLAASIALRQEWILL IG VYV ETSS
 Sbjct: 208 VIMIGALLGFFVFNHKAQVFMGDVGSALGAMLAASIALRQEWILLFIGFVYVETSS 267

50 Query: 240 VMLQVSYFKYTKKKYGEGRIFRMTFFHHHLELGGSGKGNKWEQVDAFLWGVGSLAS 299
 VMLQV+YFKYTKKK G G+RIFRMTFFHHHLELGG+SGKG KWSEW+VDAFLW +G S
 Sbjct: 268 VMLQVAYFKYTKKKTGVGKRIFRMTFFHHHLELGGVSGKGNKWEWKVDAFLWAGIFMS 327

55 Query: 300 LLVLAILYV 308
 + LAILY+
 Sbjct: 328 AITLAILYL 336

An alignment of the GAS and GBS proteins is shown below.

55 Identities = 244/309 (78%), Positives = 273/309 (87%), Gaps = 1/309 (0%)

Query: 28 LKKIGGQQMHEDVKQHLAKAGTPTMGGTVFLVALLVSLIFSIIILSKENSGNLGATFGIL 87
 LKKIGGQQMHEDVKQHLAKAGTPTMGGTVFL+VA VSL+ S+ S+N+ +L GIL
 60 Sbjct: 1 LKKIGGQQMHEDVKQHLAKAGTPTMGGTVFLVAVSLVSLF-SIKNTQSLALISGIL 59

Query: 88 SVVLIYGIIGFLDDFLKIFKQINEGLTPKQKMSLQLIAGLIFYFVHVLPSTGSAINIFGF 147
 S+V+YGIIGFLDDFLKIFKQINEGLT KQK++LQL+ GL+FYF+HV PSG S+IN+FG+
 Sbjct: 60 SIVVIYGIIGFLDDFLKIFKQINEGLTAKQKLALQLVGGLMFYFLHVSPSGISSINVFGY 119

-1049-

Query: 148 YLEVGYLYAFFVLFVWVGFSNAVNLTGIDGLASISVVISLITYGIIAYNQTFDILLII 207
 L +G Y FFVLFVWVGFSNAVNLTGIDGLASISVVISL+TYG+IAY Q+QFD+LL+I
 Sbjct: 120 QLPLGIFYLFFVLFVWVGFSNAVNLTGIDGLASISVVISLVTYGVIAVYVQSQFDVLLLI 179

5 Query: 208 VIMIGALLGFFVFNHHPAKVFMGDVGSALGAMLAAISIALRQEWILLFIGFVVVFETSS 267
 MIGALLGFF FNHHPAKVFMGDVGSALGAMLAAISIALRQEWILL IG VYV ETSS
 Sbjct: 180 GAMIGALLGFFCFNHHPAKVFMGDVGSALGAMLAAISIALRQEWILLIIGIVVLETSS 239

10 Query: 268 VMLQVAYFKYTKKKTKGVGKRIFRMTFPFHHHLELGGVSGKGNKWEKVD AFLWAIGIFMS 327
 VMLQV+YFKYTKKK G G+RIFRMTFPFHHHLELGG+SGKG KWSEW+VDAFLW +G S
 Sbjct: 240 VMLQVSIFYTKKKYGEGRIFRMTFPFHHHLELGGVSGKGNKWEKVD AFLWVGWGLAS 299

Query: 328 AITLAILYL 336
 + LAILY+
 15 Sbjct: 300 LLVLAILYV 308

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 956

20 A DNA sequence (GBSx1014) was identified in *S. galactiae* <SEQ ID 2913> which encodes the amino acid sequence <SEQ ID 2914>. This protein is predicted to be autoaggregation-mediating protein (deaD). Analysis of this protein sequence reveals the following:

Possible site: 29
 >>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3018(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

30 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB14444 GB:Z99116 similar to ATP-dependent RNA helicase
 [Bacillus subtilis]

35 Identities = 215/436 (49%), Positives = 310/436 (70%), Gaps = 5/436 (1%)

Query: 3 FKDFNFKPYIQRALDELKFDVPTDVQAKLIPVVRSGRDLVGESKTGSGKTHFLLPPIFEK 62
 F+ + KP+I A+ L F +PTD+Q +LIP V ++G+S+TG+GKTH +LLP+ K
 Sbjct: 6 FELYELKPFIIIDAVHRLGFYEPTDIQKRLIPAVLKKEVIGSQGTGKTAYLLPLLNK 65

40 Query: 63 LDESSDDVQVVITAPSRELGTQIYQATKQIAEHSE-QEIRVVNYVGGTDKLRQIEKLKVS 121
 +D + D VQVVITAP+REL QIYQ +I + E +IR ++GGTDK + I+KLK+
 Sbjct: 66 IDPAKDVVQVVITAPTRELANQIYQEALKITQGBEGSQIRSKCFIGGTDKQKSIDKLI- 124

45 Query: 122 QPHIVIGTPGRIYDLVKSGLAIHKAHTFVVEADMTLDMGFLDTVDKIAGSLPKDVQIL 181
 QPH+V+GTPGRI DL+K L++HKA + V+DEAD+ LDMGFL VD I +P+D+Q+L
 Sbjct: 125 QPHLVGTPGRIADLIKEQALSVHKAESLVIDEADLMLDMGFLADVDTIGSRMPEDLQML 184

50 Query: 182 VFSATIPQKLQPFLLKKYLTPNPMKIKTATVIADTIDNWLSTKGRDKNAQILELSKLMQ 241
 VFSATIP+KL+PFLKKY+ NP ++ V A I++ L+ +K RDK+ + ++ +
 Sbjct: 185 VFSATIPKLPFLKKYMNPKYAHVEPKQVTAAKIEHILIPSKHRDKDKLLFDIMSHLN 244

55 Query: 242 PYLAMIFVNTKERADELHSHYSSNGLKVAKHGGIAPRERKRIMNQVKNLEFEYIVATDL 301
 PYL ++F NTK AD + YL+ G+K+ +HGG+ PRERK++M Q+ +LEF YI+ATDL
 Sbjct: 245 PYLGIVFANTKNTADHIAQYLTGKGMKIGLLHGGLT'PRERKKVMKQINDLEFYIATDL 304

Query: 302 AARGIDIEGVSHVINDAIPQDLFFVHRVGRGTGRNLSGTAITLYQPSDDSDIRELEKLG 361
 AARGIDI+GVSHVIN +P DL F+VHRVGRGT R G SG A+T+Y+ +D+ + LEK+G
 Sbjct: 305 AARGIDIKGVSHVINYELPDDLDFFVHRVGRGTARAGSSGQAMTIYELTDEDALVRLEKMG 364

60 Query: 362 INFIPKVIKNGEFQDTYDRDRNRNREKSYQKLDTEMIGLVKKKKKKIKPGYKKKIQWKVD 421
 I F ++ GE++ DR RR R+K+ + D E+ + KK KK+KPGYKKK+ ++++

-1050-

Sbjct: 365 IEFEYLELEKGEWKKGDDRQRKRKRKTPNEAD-EIAHRLVKKPKVKPGYKKKMSYEME 423

Query: 422 EKRRKERRASNRKGR 437

+ ++K+RR N++K R

5 Sbjct: 424 KIKKKQRR--NQSKKR 437

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2915> which encodes the amino acid sequence <SEQ ID 2916>. Analysis of this protein sequence reveals the following:

Possible site: 54

10 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2315(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

15 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 382/447 (85%), Positives = 420/447 (93%)

20 Query: 1 MSFKDFNFKPYIQRALDELKFVDPTDVQAKLIPVVRSGRDLVGESKTGSGKTHTFLLPIF 60
MSFKD++FK Y+Q+AL+E+ FV+PT+VQ +LIP+V SGRDLVGESKTGSGKTHTFLLPIF
Sbjct: 1 MSFKDYHFKQYVQQALEEIGFVNPTFVQKRLIPVNSGRDLVGESKTGSGKTHTFLLPIF 60

25 Query: 61 EKLEDESSDDVQVVITAPSRELGTQIYQATKQIAEHSEQEIRVVNYVGGTDKLRQIEKLV 120
EKLEDE+ +VQVVITAPSREL TQI+ A KQIA+H ++EIR+ NYVGGTDKLRQIEKLV
Sbjct: 61 EKLEDEAKAEVQVVITAPSRELATQIFDACKQIAKHFQEEIRLANVYVGGTDKLRQIEKLV 120

30 Query: 121 SQPHIVIGTPGRIYDLVKSGDLAIHKAHTFVVDADMTLDMGFLDTVDKIASLPKDVQI 180
SQPHIVIGTPGRIYDLVKSGDLAIHKA TFVVDADMT+DMGFLDTVDKIA SLPK VQI
Sbjct: 121 SQPHIVIGTPGRIYDLVKSGDLAIHKATTFVVDADMTMDMGFLDTVDKIAASLPKSVQI 180

35 Query: 181 LVFSATIPQKLQFFLKLYLTNPVMEKIKTATVIADTIDNWLSTKGRDKNAQILELSKLM 240
LVFSATIPQKLQFFLKLYLTNPV+E+IKT TVIADTIDNWL+STKGRDKN Q+LE+ K M
Sbjct: 181 LVFSATIPQKLQFFLKLYLTNPVIEQIKTKTVIADTIDNWLSTKGRDKNQLEILKTM 240

40 Query: 241 QPYLAMIFVNTKERADELHLSYSSNGLKVAKIHGGIAPRERKRIMNQVKNLEFEYIVATD 300
QPY+AM+FVNTKERAD+LH++L++NGLKVAKIHGGI PRERKRIMNQV L+FEYIVATD
Sbjct: 241 QPYMAMLFVNTKERADDLHAFLTANGLKVAKIHGGIPRERKRIMNQVKLDFEYIVATD 300

45 Query: 301 LAARGIDIEGVSHVINDAIPQDLSFFVHRVGRGTRNGLSGTAITLYQPSDDSDIRELEKL 360
LAARGIDIEGVSHVINDAIPQDLSFFVHRVGRGTRNG++GTAITLYQPSDDSDI+ELEK+
Sbjct: 301 LAARGIDIEGVSHVINDAIPQDLSFFVHRVGRGTRNGMAGTAITLYQPSDDSDIKELEKM 360

50 Query: 361 GINFIPKVIKNGEFQDTYDRRRNNREKSYQKLDTEMIGLVKKKKKKIKPGYKKKIQWKV 420
GI F PKV+KNGEFQDTYDRRR NREK+YQKLDTEMIGLVKKKKKK+KPGYKKKIQW V
Sbjct: 361 GIAFTPKVLKNGEFQDTYDRRRQREKAYQKLDTEMIGLVKKKKKKVKPGYKKKIQWAV 420

Query: 421 DEKRRKERRASNRKGRRAERKAKKQSF 447
DEKRRKERRA NRAKGRRAERKAKKQ F
Sbjct: 421 DEKRRKERRAENRAKGRRAERKAKKQHF 447

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 957

55 A DNA sequence (GBSx1015) was identified in *S.agalactiae* <SEQ ID 2917> which encodes the amino acid sequence <SEQ ID 2918>. This protein is predicted to be unnamed protein product. Analysis of this protein sequence reveals the following:

Possible site: 19

>>> Seems to have a cleavable N-term signal seq.

-1052-

230 240 250 260

SEQ ID 8694 (GBS8) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 2 (lane 5; MW 31kDa), Figure 63 (lane 2; MW 31.3kDa), Figure 66 (lane 2 & 3; MW 31kDa), in Figure 178 (lane 2; MW 31kDa), in Figure 179 (lane 3 & 4; MW 31kDa) and in Figure 180 (lane 3; MW 31kDa). It was also expressed in *E.coli* as a GST-fusion product, with SDS-PAGE shown in Figure 66 (lanes 4 & 5; MW 56kDa) and in Figure 180 (lanes 4 & 5; MW 55kDa).

GBS8-His was purified as shown in Figures 189 (lane 7), 211 (lane 3), 228 (lanes 4-5) and 230 (lanes 3-6). Purified GBS8-GST is shown in Figure 209, lane 6.

The GBS8-His fusion product was purified (Figure 90A) and used to immunise mice (lane 2 product; 12.9µg/mouse). The resulting antiserum was used for Western blot (Figure 90B), FACS (Figure 90C), and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 958

A DNA sequence (GBSx1016) was identified in *S.agalactiae* <SEQ ID 2921> which encodes the amino acid sequence <SEQ ID 2922>. Analysis of this protein sequence reveals the following:

```

Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3991(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 959

A DNA sequence (GBSx1017) was identified in *S.agalactiae* <SEQ ID 2923> which encodes the amino acid sequence <SEQ ID 2924>. This protein is predicted to be probable amino-acid abc transporter permease protein in *idh-deor* inter. Analysis of this protein sequence reveals the following:

```

Possible site: 56
>>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -11.62    Transmembrane    50 - 66 ( 41 - 74)
      INTEGRAL    Likelihood = -0.90     Transmembrane    226 - 242 ( 226 - 242)
      INTEGRAL    Likelihood = -0.53     Transmembrane    80 - 96 ( 80 - 96)

----- Final Results -----
      bacterial membrane --- Certainty=0.5649(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

-1053-

>GP:CAB15985 GB:Z99124 similar to amino acid ABC transporter
(permease) [Bacillus subtilis]

Identities = 90/224 (40%), Positives = 137/224 (60%), Gaps = 10/224 (4%)

5 Query: 28 WKAVIDAIPISILERLPITLLLTVAGALEGLILALIFAVVKINRVKILYPIQALFVSFLRG 87
W+ ++ A P++++ LPITL + +A +F +I LI A++ N++ +L+ + L++SF RG
Sbjct: 6 WEFMISAFPTLIQALPITLFMAIAAMIFAIIGGLILALITKNKIPVLHQLSKLYISFFRG 65

10 Query: 88 TPILVQLMLSYYGIPFLKFLNQKYGFDFWNINAI PASVFATAFNEAAYTSETIRAAI 147
P LVQL L YYG+P +++ + A AI + AAY +E RAA+
Sbjct: 66 VPTLVQLFLIYYGLPQLFPEMSK-----MTALTAAIIGLSLKNAAYLAEIFRAAL 115

15 Query: 148 LSVDQGEIEAARSLGMTSAQVYRRVVIIPNAAVVATPTLINTLIGLTKGTSIAFNAGIVEM 207
SVD G++EA S+GMT Q YRR+I+P A A P NT IGL K TSLAF G++EM
Sbjct: 116 NSVDDQGLEACLSVGMTKFQAYRRRIILPQAIRNAIPATGNTFIGLLKETSLAFTLGVMEM 175

20 Query: 208 FAQAQIMGSDYRYFERYISVALVYWA VSFLEQLGNAIERKMA 251
FAQ ++ + +YFE Y++VA+VYW ++ + L + ER M+
Sbjct: 176 FAQGKMYASGNLKYFETYLAIVAIVYVWLTIIYSILQDLFERAMS 219

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2925> which encodes the amino acid sequence <SEQ ID 2926>. Analysis of this protein sequence reveals the following:

Possible site: 43

>>> Seems to have no N-terminal signal sequence

25 INTEGRAL Likelihood = -7.27 Transmembrane 80 - 96 (74 - 104)
INTEGRAL Likelihood = -1.06 Transmembrane 207 - 223 (207 - 223)
INTEGRAL Likelihood = -0.90 Transmembrane 110 - 126 (110 - 126)

----- Final Results -----

30 bacterial membrane --- Certainty=0.3909(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related sequence was also identified in GAS <SEQ ID 9167> which encodes the amino acid sequence <SEQ ID 9168>. Analysis of this protein sequence reveals the following:

Possible site: 60

>>> Seems to have no N-terminal signal sequence

40 INTEGRAL Likelihood = -7.27 Transmembrane 50 - 66 (44 - 74)
INTEGRAL Likelihood = -1.06 Transmembrane 177 - 193 (177 - 193)
INTEGRAL Likelihood = -0.90 Transmembrane 80 - 96 (80 - 96)

----- Final Results -----

45 bacterial membrane --- Certainty=0.391(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 212/267 (79%), Positives = 238/267 (88%)

50 Query: 1 MNQFILTGGWSWYNNLVSQVPAGKLFWSKAVLDAIPISILERLPITLLLTVAGALEGLILA 60
M LT GW++Y+ L+S +P GKLFWS AV DAIP+I++RLPITL LT++GA FGL+IA
Sbjct: 31 MTSVFLTSGWAFYDYLISFIPHGKLFWSHAVFDAIPNIIQRLPITLGLTLSGATFGLVLA 90

55 Query: 61 LIFAVVKINRVKILYPIQALFVSFLRGTPILVQLMLSYYGIPFLKFLNQKYGFDFWNINA 120
LIFA+VKIN+VK+LYPIQA+FVSFLRGTPILVQLML+YYGIPFLKFLNQKYGFDFWN+NA
Sbjct: 91 LIFALVKINKVKLLYPIQAFVSFLRGTPILVQLMLTYYGIPFLKFLNQKYGFDFWNVNA 150

60 Query: 121 IPASVFATAFNEAAYTSETIRAAILSVDQGEIEAARSLGMTSAQVYRRVVIIPNAAVV 180
IPAS+FAITAFNEAAY SETIRAAILSVD GEIEAA+SLGMTS QVYRRVVIIPNA VV
Sbjct: 151 IPASIFAITAFNEAAYASETIRAAILSVDGTGEIEAAKSLGMTSVQVYRRVVIIPNATVV 210

Query: 181 ATPTLINTLIGLTKGTSIAFNAGIVEMFAQAQIMGSDYRYFERYISVALVYWA VSFLE 240
A PTLIN LIGLTKGTSIAFNAGIVEMFAQAQI+GSDYRYFERYISVALVYW++S L+E

-1054-

Sbjct: 211 AIPTLINGLIGLTKGTSLAFNAGIVEMFAQAQILGGSDYRYFERYISVALVYWSISILME 270

Query: 241 QLGNAIERKMAIKAPRHLTDEIPGGVR 267

Q+G IE KMAIKAP +E G +R

5 Sbjct: 271 QVGRLIENKMAIKAPEQARNEKLGEIR 297

There is also homology to SEQ ID 4794.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

10 Example 960

A DNA sequence (GBSx1018) was identified in *S.galactiae* <SEQ ID 2927> which encodes the amino acid sequence <SEQ ID 2928>. This protein is predicted to be amino acid ABC transporter, ATP-binding protein. Analysis of this protein sequence reveals the following:

Possible site: 46

15 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3205(Affirmative) < succ>

20 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC00329 GB:AF008220 putative amino acid transporter [Bacillus subtilis]
Identities = 121/247 (48%), Positives = 176/247 (70%)

25 Query: 1 MIKLRQLTKSFGQKVLDDLDIEKGQVVALVGASGAGKSTFLRSMNYLEPDYGTIEI 60

MI+++ + K F VL ++L + KG+VV ++G SG+GK+TFLR +N LE PD G I I

Sbjct: 1 MIEIKNIHKQFGIHHVLKGINLTVRKGEVVTIIGPSGSGKTTFLRCINLLERPDEGIISI 60

30 Query: 61 DDFKVFPSISKDDILTLLRRKLAMVFQQFNLFFERRTALDNVKEGLKIVKKMSDQEATRIA 120

D ++ + SK ++ LR++ AMVFQQ++LF +T ++NV EGL I +KM Q+A +A

Sbjct: 61 HDKVINCRRFPSKKEVHWLRKQTAMVFQQYHLFAHKTIVENVMEGLTIARKMRKQDAYAVA 120

35 Query: 121 RDELAKVGLADREKYYPRLSGGQKQORVALARALAMKPDVLLDEPTSAIDPELVGEVEK 180

+EL KVGL D+ YP LSGGQKQRV +ARALA+ PDVLL DEPT+ALDPELVGEV +

Sbjct: 121 ENELRKVGLQDKLNAYPSQLSGGQKQRVGIARALAIHPDVLLFDEPTAALDPELVGEVLE 180

40 Query: 181 SIADAAKQGQTMVLVSHDMNFVYQVADKVLFLKGRILESGTPEQLFNHPLEERTKEFFA 240

+ + K G TM++V+H+M F +V+D+V+F+++G I+E GTPE++F H ++RT++F

Sbjct: 181 VMLEIVKTGATMIVVTHEMEFARRVSDQVVFMDQVIVEQGTPEEVFRHTKKDRTRQFLR 240

Query: 241 SYNKSYL 247

+ YL

45 Sbjct: 241 RVSPEYL 247

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2929> which encodes the amino acid sequence <SEQ ID 2930>. Analysis of this protein sequence reveals the following:

Possible site: 13

50 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1840(Affirmative) < succ>

55 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 199/247 (80%), Positives = 229/247 (92%)

-1055-

Query: 1 MIKLRQLTKSFSGQKVLDDKLDLDIEKGQVVALVGASGAGKSTFLRSMNYLEEPDYGTIEI 60
 MI +R L+K+FSGQKVLDD LLDIEKGQV+ALVGASGAGKSTFLRS+NYLE+PD G+I I
 Sbjct: 2 MITIRNLSKTFSGQKVLDSLALDIEKGQVIALVGASGAGKSTFLRSINYLEKPDGSGSISI 61

5 Query: 61 DDFKVDKFSISKDDILTLRRKIAMVFQQFNLFERRTALDNVKEGLKIVKKMSDQEATRIA 120
 DF VDF++I+ + +L LRRKIAMVFQQFNLFERRTAL+NVKEGLK+VKK+SDQEAT++A
 Sbjct: 62 GDFTVDFETITTEQVLILRRKIAMVFQQFNLFERRTALENVKEGLKVKKLSDQEATKLA 121

10 Query: 121 RDELAKVGLADREKYYPRHLSGGQKQORVALARALAMKPDVLLLDDEPTSALDPELVGEVEK 180
 + ELAKVGLADR+ +YPRHLSGGQKQORVALARALAMKPDVLLLDDEPTSALDPELVGEVEK
 Sbjct: 122 QAELAKVGLADRKHHYPRHLSGGQKQORVALARALAMKPDVLLLDDEPTSALDPELVGEVEK 181

15 Query: 181 SIADAAKQGQTMVLVSHDMNFVYQVADKVLFLBKGRILESGTPEQLFNHPLEERTKEFFA 240
 SI DAAK GQTMVLVSHDMNFVYQVAD+VLFL++G+ILE GTPE++F HP +ERTKEFFA
 Sbjct: 182 SITDAKSGQTMVLVSHDMNFVYQVADRVLFDDQGKILEQGTPEEVFRHPQKERTKEFFA 241

Query: 241 SYNKSYL 247
 SY+K+Y+
 20 Sbjct: 242 SYSKTYI 248

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 961

25 A DNA sequence (GBSx1019) was identified in *S.agalactiae* <SEQ ID 2931> which encodes the amino acid sequence <SEQ ID 2932>. Analysis of this protein sequence reveals the following:

Possible site: 17
 >>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0831(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

35 The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB07290 GB:AP001519 thioredoxin reductase (NADPH) [Bacillus halodurans]
 Identities = 173/302 (57%), Positives = 234/302 (77%)

40 Query: 1 MYDTLIIGSGPGGTAALYAARSLKVLIEQAGPGQMNTABIENTYPGYDHISGPPELS 60
 +YD +I G+GP GMTAA+Y +R+NL ++E+G PGGQM NT ++ENYPG+DHI GPPELS
 Sbjct: 7 VYDVVIAGAGPAGMTAAVYTSRANLSTVMVERGVPGQMANTEDVENYPGFDHILGPPELS 66

45 Query: 61 MKMYEPLKFEVEHIYQVQVENDGDVVRVITEDESYEAKTVILATGAKNSLLGVPGE 120
 KM+E +KF E+ YG ++ + + GD+K V ++ Y+A+ VI+ATGA+ LGVPGE+
 Sbjct: 67 TKMFEHAKKFGAEYAYGDIKEIIDQDGLKLVKAGNKEYKARAVIVATGAEYKKGVPGEK 126

50 Query: 121 EYTSRGVSYCAVCDGAFFRDQDLLVGGGDSAVEEAVFLTQFAKSVTIIHRRDQLRAQKV 180
 E + RGVSYCAVCDGAFF+ ++L+VVGGSASAVEEAV+LT+FA VTIIHRRDQLRAQK+
 Sbjct: 127 ELSRGVSYCAVCDGAFFKGKELVVVGGSASAVEEAVYLTRFASKVTIIHRRDQLRAQKI 186

55 Query: 181 LQDRAFANEKIKFVWDSVVKIKGNEIKVSGVTVENLKTGEISEMTFGGVFIYVGLKPHS 240
 LQ RAF N+KI+F+WD VVK+I G + KVS VT+E+ KTGE + GVFIY+G+ P +
 Sbjct: 187 LQQRATNDKIEFIWDHVVKQINGTDGKVSSVTIEHAKTGEQQDFKTDGVFIYIGMLPLN 246

60 Query: 241 SMVSELGITDETGWVLTDTNMKTSIPGLYAIGDVROKDLRQIATAVGEGAIAGQGVYNYI 300
 V L I ++ G+++T+ M+TS+PG++A GDVR+K LRQI TA G+G++A Q V +YI
 Sbjct: 247 EAVKLNILNDEGYIVTNEEMETSVPGIFAAGDVREKSLRQIVTATGDGSLAAQNVQHYI 306

Query: 301 TE 302
 E
 Sbjct: 307 EE 308

-1056-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2933> which encodes the amino acid sequence <SEQ ID 2934>. Analysis of this protein sequence reveals the following:

Possible site: 28

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.0386(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 236/300 (78%), Positives = 273/300 (90%)

Query: 1 MYDTLIIGSGPGMTAALYAARSNLKVGLIEQGAPGGQMNNTAEIENYPGYDHISGPELS 60

MYDTLIIGSGP GMTAALYAARSNL V +IEQGAPGGQMNNT +IENYPGYDHISGP EL+

Sbjct: 1 MYDTLIIGSGPAGMTAALYAARSNLSVAIEQGAPGGQMNNTFDIENYPGYDHISGP ELA 60

Query: 61 MKMYEPLKFEVEHIYQIVQRVENDGDVVRVITEDES YEAKTVILATGAKNSLLGVP GEE 120

MKMYEPLK F VE+IYQIVQ++EN GD K V+TED SYEAKTVI+ATGAK +LGVP GEE

Sbjct: 61 MKMYEPLKFNVENIYQIVQKIENFGDYKCVLTEDAS YEAKTVI IATGAKYRVLGVP GEE 120

Query: 121 EYTSRGVSYCAVCDGAFFRDQDLLVVG GDSAVEEAVFLTQFAKSVTIIHRRDQLRAQKV 180

YTSRGVSYCAVCDGAFFRDQDLLVVG GDSAVEE++LTQFAK VT++HRRDQLRAQK+

Sbjct: 121 YTSRGVSYCAVCDGAFFRDQDLLVVG GDSAVEEAIYLTQFAKVTIVHRRDQLRAQKI 180

Query: 181 LQDRAFANEKIKFVWDSVVK EIKGNEIKVSGVTVENLKTGEISEMTFGGVFIYVGLKPHS 240

LQDRAFAN+K+ F+WDSVVK E I+GN+IKVS V +EN+KTG++++ FGGVFIYVG+ P +

Sbjct: 181 LQDRAFANDKVDFIWDSVVK EIQGNDIKVSNVLIENVKTGQVTDHAFGGVFIYVGMNFVT 240

Query: 241 SMVSELGITDETGWVLTDTNMKTSIPGLYAIGDVRQKDLRQIATAVGE GAIAGQGVNYI 300

MV +L ITD GW++TD +M+TSIPG++AIGDVRQKDLRQI TAVG+GAIAGQGVY+Y+

Sbjct: 241 GMVKDLEITDSEGWITDDHMRTSIPGIFAIGDVRQKDLRQIT TAVGDGAIAGQGVYHYL 300

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 962

A DNA sequence (GBSx1020) was identified in *S.galactiae* <SEQ ID 2935> which encodes the amino acid sequence <SEQ ID 2936>. Analysis of this protein sequence reveals the following:

Possible site: 35

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3626(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB15163 GB:Z99120 similar to nicotinate

phosphoribosyltransferase [Bacillus subtilis]

Identities = 309/476 (64%), Positives = 384/476 (79%), Gaps = 2/476 (0%)

Query: 2 YKDDSLTLHTDLYQINMMQVYFNKG IHNKRAVF EAYFRKVPFENG YAVFAGLERIVRYLE 61

+KDDSL+LHTDLYQINM + Y+ GIH K+A+FE +FR++PFENG YAVFAGLE+ + YLE

Sbjct: 6 FKDDSLSLHTDLYQINMAET YWRDGIHEKKAIFELFFRRLPFENG YAVFAGLEKAIEYLE 65

Query: 62 NLSFSDSDLSYLE-ELGYPEEF LDYLNKRLMELTVKSAKEGDLVFANEPLVQIEGPLAQC 120

N F+DSDLSYL+ ELGY E+F++YL+ L ++ S KEG+LVF NEP++++E PL +

Sbjct: 66 NFKFTSDSDLSYLQDELGYHEDFIEYLRGLSFTGSLYSMKEGELVFNNEPIMRVEAPLVEA 125

-1057-

Query: 121 QLVETAILNIINYQTLVATKAARIRSVIEDEPLLEFGTRRAQEMDAAIWGTRAAIIGGAN 180
 QL+ETA+LNI+NYQTL+ATKAARI+ VI DE LEFGTRRA EMDAA+WG RAA+IGG +
 Sbjct: 126 QLIETALLNIVNYQTLIATKAARIKGVIGDEVALEFGTRRAHEMDAAMWGARAALIGGFS 185

5 Query: 181 ATSNVRAGKIFNIPVSGTHAHALVQTYGDDYQAFKAYAETHKDCVFLVDYDYLTVRGVFN 240
 ATSNVRAGK FNIIPVSGTHAHALVQ Y D+Y AFK YAETHKDCVFLVDYDYLTVR G+PN
 Sbjct: 186 ATSNVRAGKRFNIPVSGTHAHALVQAYRDEYTAFFKKAETHKDCVFLVDYDYLTVRSGMPN 245

10 Query: 241 AIRVAKEMGEKINFLGVRLDSGDLAYLSKKVRQQLDDAGFPNAKIYASNDLIDENTILNLK 300
 AIRVAKE G++INF+G+RLDSGDLAYLSKK R+ LD+AGF +AK+ AS+DLDE+TI+NLK
 Sbjct: 246 AIRVAKEFGDRINFIGIRLDSGDLAYLSKKARMLDEAGFTDAKVIASSDLDEHTIMNLK 305

15 Query: 301 MQKAKIDVWVGVTGLITAYDQFALGAVYKIVSIETDAGSMRDTIKLSNNAEKVSTPGKKQ 360
 Q A+IDVWVGVTGLITAYDQFALGAVYK+V+IE D G M DTIK+S+N EKV+TPG+K+
 Sbjct: 306 AQGARIDVWVGVTGLITAYDQFALGAVYKLVAIEED-GKMVDTIKISSNPEKVTPGRKK 364

20 Query: 361 VWRITSRAKGKSEGDYITFADTDVTQLDEIEMFHPTTYTYINKTVRDFDAVPLLDIFDKG 420
 V+RI +++ SEGDI D V + MFHP +T+I+K V +F A L IF+KG
 Sbjct: 365 VYRIINQSNHSEGDYIALYDEQVNDQKRLRMFHPVHTFISKVTFNIFYAKDLHELIFDKG 424

25 Query: 421 KLVYQLPSLQEIQEYGRKEFDQLWDEYKRVLPQDYPVDLARDVWQNKMDLIDRIR 476
 L YQ P + +IQ+Y + LW+EYKR+ P++YPVDL+ D W NKM I ++
 Sbjct: 425 ILCYQNPEISDIQQYVQDNLSELLWEEYKRISKPEEYPVDLSEDCWSNKMQRHEVK 480

25 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2937> which encodes the amino acid sequence <SEQ ID 2938>. Analysis of this protein sequence reveals the following:

Possible site: 32
 >>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3192(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

35 An alignment of the GAS and GBS proteins is shown below.

Identities = 409/484 (84%), Positives = 446/484 (91%)

Query: 1 MYKDDSLTLHTDLYQINMMQVYFNKGIIHNKRAVFEAYFRKVPFENGYAVFAGLERIVRYL 60
 MYKDDSLTLHTDLYQINMMQVYF +GIHN+ AVFE YFRK PF NGYAVFAGL+R+V YL
 40 Sbjct: 1 MYKDDSLTLHTDLYQINMMQVYFQGIHNRHAVFEVYFRKEPFNNGYAVFAGLQRMVEYL 60

Query: 61 ENLSFSDSDLSYLEELGYPEEFLDYLNKLMELTVKSAKEGDLVFANEPLVQIEGPLAQ 120
 E FS++DL+YLEELGYPE FL YLK L++ELT++SAKEGDLVFANEPLVQ+EGPL QC
 Sbjct: 61 EQFQFSETDLAYLEELGYPENFLTLYKELRLLELTIRSAKEGDLVFANEPLVQIEGPLQC 120

45 Query: 121 QLVETAILNIINYQTLVATKAARIRSVIEDEPLLEFGTRRAQEMDAAIWGTRAAIIGGAN 180
 QLVETA+LNI+N+QTL+ATKAARIRSVIEDEPLLEFGTRRAQE+DAAIWGTRAA+IGGA+
 Sbjct: 121 QLVETALLNIVNFQTLIATKAARIRSVIEDEPLLEFGTRRAQELDAAIWGTRAAIIGGAD 180

50 Query: 181 ATSNVRAGKIFNIPVSGTHAHALVQTYGDDYQAFKAYAETHKDCVFLVDYDYLTVRGVFN 240
 ATSNVRAGK F+IPVSGTHAHALVQ YG+DY AF AYA+THKDCVFLVDYDYLTVR GVP
 Sbjct: 181 ATSNVRAGKRFDIPVSGTHAHALVQAYGNDYDAFMAYAKTHKDCVFLVDYDYLTVRGVFT 240

55 Query: 241 AIRVAKEMGEKINFLGVRLDSGDLAYLSKKVRQQLDDAGFPNAKIYASNDLIDENTILNLK 300
 AIRVAKEMG+KINFLGVRLDSGDLAYLSK VRQQLDDAGF AKIYASNDLIDENTILNLK
 Sbjct: 241 AIRVAKEMGDKINFLGVRLDSGDLAYLSKTVRQQLDDAGFTEAKIYASNDLIDENTILNLK 300

60 Query: 301 MQKAKIDVWVGVTGLITAYDQFALGAVYKIVSIETDAGSMRDTIKLSNNAEKVSTPGKKQ 360
 MQKAKIDVWVGVTGLITAYDQFALGAVYKIVSIE + GSMRDTIKLSNNAEKVSTPGKKQ
 Sbjct: 301 MQKAKIDVWVGVTGLITAYDQFALGAVYKIVSIEQEDGSMRDTIKLSNNAEKVSTPGKKQ 360

65 Query: 361 VWRITSRAKGKSEGDYITFADTDVTQLDEIEMFHPTTYTYINKTVRDFDAVPLLDIFDKG 420
 VWRITSR KGKSEGDYITF D +V +L EIEFMHPTTYTYI KTV++FDA+PLLDIF KG
 Sbjct: 361 VWRITSREKGKSEGDYITFTDINVNELTEIEMFHPTTYTYIKTVKEFDAIPLLDIFVK 420

-1058-

Query: 421 KLVYQLPSLQEIQEYGRKEFDQLWDEYKRVLPQDYPVDLARDVWQNKMDLIDIRKEAL 480
 +LVYQLP+L EI+ Y +KEFD+LWDEYKRVLPQDYPVDLARDVWQNM LID IRK+A
 Sbjct: 421 ELVYQLPTLAEIKAYAKKEFDKLWDEYKRVLPQDYPVDLARDVWQNMALIDNIRKDAY 480

5 Query: 481 AKGE 484
 K E
 Sbjct: 481 GKSE 484

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 963

A DNA sequence (GBSx1021) was identified in *S.galactiae* <SEQ ID 2939> which encodes the amino acid sequence <SEQ ID 2940>. Analysis of this protein sequence reveals the following:

Possible site: 56
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2744(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC74810 GB:AB000269 NAD synthetase, prefers NH3 over glutamine
 [Escherichia coli K12]
 Identities = 173/274 (63%), Positives = 214/274 (77%), Gaps = 1/274 (0%)

Query: 1 MTLQDQIIKELGVKPVINPSQEIIRRSVEFLKDYLLKHSFLKTYVLGISGGQDSTLAGRIA 60
 MTLQ QIIK LG KP IN +EIRRSV+FLK YL + F+K+ VLGISGGQDSTLAG+L
 Sbjct: 1 MTLQQQIIKALGAKPQINAEIEIRRSVDFLKSYLQTYPFIKSLVLGISGGQDSTLAGKLC 60

Query: 61 QLAVEELRADTG-ENYQFIAIRLPYGIQADEEDAQKALDFIKPDIALTINIKEAVDGGQVR 119
 Q+A+ ELR +TG E+ QFIA+RLPYG+QADE+D Q A+ FI+PD LT+NIK AV +
 Sbjct: 61 QMAINELRLETGNESLQFIAVRLPYGVQADEQDCQDAIAFIQPDRLTVNKGAVLASEQ 120

Query: 120 ALNAGVEITDFNKGNIKARQRMISQYAVAGQYAGAVIGTDHAAENITGFFTKFGDGGAD 179
 AL AG+E++DF +GN KAR+RM +QY++AG +G V+GTDHAAE ITGFFTK+GDGG D
 Sbjct: 121 ALREAGIELSDFVRGNEKARERMAQYSIAGMTSGVVVGTDHAAEAITGFFTKYGDGGTD 180

Query: 180 LPLPLFRLNKSGQKQLLAELGADKALYEKIPTADLEENKPGIADLEIALGVITYQEIDAYLEG 239
 + PL+RLNK QGKQLLA L + LY+K PTADLE+++P + DE+ALGVTY ID YLEG
 Sbjct: 181 INPLYRLNKRQGGKQLLAALACPEHLYKKAPTADLEDDRPSLPDEVALGVITYDNIDDYLEG 240

Query: 240 KVSDDKSRGIENWWYKQHKRHLPTITFDDFWK 273
 K V + IENW+ K +HKR PIT+FDDEFWK
 Sbjct: 241 KNPQQVARTIENWYLKTEHKRRPPTITVDDFWK 274

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2941> which encodes the amino acid sequence <SEQ ID 2942>. Analysis of this protein sequence reveals the following:

Possible site: 18
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3482(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 213/274 (77%), Positives = 242/274 (87%), Gaps = 1/274 (0%)

-1059-

Query: 1 MTLQDQIIKELGVKPVINPSQEIIRSVFELKDYLLKHSFLKTYVLGISGGQDSTLAGRLA 60
 MTLQ++II++LGVK I+P +EIR++V+FLK YL KHSFLKTYVLGISGGQDSTLAG+LA
 Sbjct: 15 MTLQEEIIRQLGVKASIDPQEEIRKAVDFLKAYLRKHSFLKTYVLGISGGQDSTLAGKLA 74

5 Query: 61 QLAVEELRADTGEN-YQFIAIRLPYGIQADEEDAQKALDFIKPDIAL/TINIKEAVDGQVR 119
 Q+A+ ELR + + YQFIA+RLPYG+QADE DAQKAL FI PD LTINIK AVDGQV
 Sbjct: 75 QMAIAELREEASDQAYQFIAVRLPYGVQADEEDAQKALAFIAPDQTLTINIKAAVDGQVE 134

10 Query: 120 ALNAAGVEITDFNKGNIKARQRMISQYAVAGQYAGAVIGTDHAAENITGFFTKFGDGGAD 179
 AL AAGVEI+DFNKGNIKARQRMISQYA+AGQ AGAVIGTDHAAENITGFFTKFGDGGAD
 Sbjct: 135 ALQAAGVEISDFNKGNIKARQRMISQYALAGQMAGAVIGTDHAAENITGFFTKFGDGGAD 194

15 Query: 180 LLPLFRLNKSQKGQLLAELGADKALYEKIPTADLEENKPGIADLALGVTYQEIDAYLEG 239
 +LPLFRLNK QGK LL LGAD ALYEK+PTADLE+ KPG+ADE+ALGVTYQ+ID YLEG
 Sbjct: 195 ILPLFRLNKRQKALKVLGADAALYEKVPTADLEDQKPLADEVALGVTYQDIDDYLEG 254

20 Query: 240 KVVSDKSRGIENWWYKGQHKRHLPTIFDDFWK 273
 K++S ++ IE WW+KGQHKRHLPTIFDDFWK
 Sbjct: 255 KLISKVAQATIEKWWHKQHKRHLPTIFDDFWK 288

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 964

A DNA sequence (GBSx1022) was identified in *Sagalactiae* <SEQ ID 2943> which encodes the amino acid sequence <SEQ ID 2944>. Analysis of this protein sequence reveals the following:

Possible site: 28
 >>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2718(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

35 >GP:CAA82960 GB:Z30315 aminopeptidase C [Streptococcus thermophilus]
 Identities = 363/444 (81%), Positives = 407/444 (90%)

40 Query: 1 MSKLTQTFTDKLFADYQANTKFSAIENAVTHNGLLKSLETRQSEIENDYVFSIDLTKDEV 60
 M+ L+ FT+KLFADY+AN K+ AIENAVTHNGLLKS+ETRQSE+END+VFSIDLTKDEV
 Sbjct: 1 MTSLSSTDFTEKLFADYEANAKYGAENAVTHNGLLKSLETRQSEVENDFVFSIDLTKDEV 60

45 Query: 61 SNQKQSGRCWMFAALNTFRHKLISDFKLENFELSQAHTFFWDKYEKSNWFMEQIIATANQ 120
 SNQK SGRRCWMFAALNTFRHKLISDFKLE+FEELSQAHTFFWDKYEKSNWF+EQIIATA+Q
 Sbjct: 61 SNQKASGRRCWMFAALNTFRHKLISDFKLESFELSQAHTFFWDKYEKSNWFLEQIIATADQ 120

50 Query: 121 ELSSRKVKFLLDVPQQDGGQWDMVVALFEKYGVVPKTVYPESVSSASRELNQYLNKLLR 180
 E+ SRKVKFLLD PQDGGQWDMVV+LFEKYGVVPK+VYPEV+SS SRELNQYLNKLLR
 Sbjct: 121 EIGSRKVKFLLDTPQQDGGQWDMVVALFEKYGVVPKSVYPEVASSNSRELNQYLNKLLR 180

55 Query: 181 QDAQILRELIAQGADGATVQNKKEELLQEIFNFLAMNLGLPPQSFDFAYRDKDNHYQSDK 240
 QDAQILR+LIA GAD A VQ KKEE LQEIFN+LIAM LGLPP+ FDFAYRDKD++Y+S+K
 Sbjct: 181 QDAQILRLIASGADQAAVQAKKEEFLQEIFNYLAMTLGLPPRQFDFAYRDKDDNYRSEK 240

60 Query: 241 NITPKAFYQKYVNLDSLSDYVSIINAPTVDKPYGQSYTVEMLGNVVGPAVKYLNLDKMRK 300
 ITP+AF++KYV L LSDYVS+INAPT DKPYG+SYTVEMLGNVVG P+V+Y+NL M RF
 Sbjct: 241 GITPRAFFEKYVGLKLSLDYVSVINAPTADKPYGKSYTVEMLGNVVGAPSVRYINLPMDRF 300

65 Query: 301 KELAIQAQMKSGETVWFGSDVGQVSNRQKGILATTTYDFNSSMDIKLSQDKAGRLDYSESL 360
 KELAIQAQMK+GE+VWFGSDVGQVS+RQKGILAT YDF +SMDI +QDKAGRLDYSESL
 Sbjct: 301 KELAIQAQMKAGESVWFGSDVGQVSDRQKGILATNVYDFTASMDINWTQDKAGRLDYSESL 360

Query: 361 MTHAMVLTVGDLDESGQPLKWKVENSWEKVKDGYFVASDAWMDEYTYQIVVRKELLTK 420

-1060-

MTHAMVLTGVLDL G+P+KWK+ENSWG+KVG+ GYFVASDAWMDEYTYQIVVRK+ LT
 Sbjct: 361 MTHAMVLTGVLDLADGKPIKWKIENSWGDKVGQKGYFVASDAWMDEYTYQIVVRKDFLTA 420

Query: 421 EELEAYNAEPITLAPWDPMGALAN 444
 EEL AY A+P LAPWDPMG+LA+
 Sbjct: 421 EELAAYEADPQVLAPWDPMGSLAS 444

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2945> which encodes the amino acid sequence <SEQ ID 2946>. Analysis of this protein sequence reveals the following:

Possible site: 14
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3002(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 369/443 (83%), Positives = 407/443 (91%)

Query: 1 MSKLTQTFTDKLFADYQANTKFSAIENAVTHNGLLSLETRQSEIENDYVFSIDLTKDEV 60
 MS LT+TFT++LFA Y+AN KFSAIENAVTHNGLLSLETRQSE++ND+VFSIDLTKD+V
 Sbjct: 1 MSALTETFTQLFAHYEANA KFSAIENAVTHNGLLSLETRQSEVDNDFVFSIDLTKDKV 60

Query: 61 SNQKSGRCWMFAALNTFRHKLISDFKLENFELSQAHTFFWDKYEKSNWFMEQIATANQ 120
 SNQK SGRCWMFAALNTFRHKLII++FKLENFELSQAHTFFWDKYEK+NWFEQ+IATA+Q
 Sbjct: 61 SNQKASGRCWMFAALNTFRHKLITEFKLENFELSQAHTFFWDKYEKANWFMEQVIATADQ 120

Query: 121 ELSSRKVKFLLDVPQQDGGQWDMVVALFEKYGVVPKTVYPESVSSASRELNQVLNKLRL 180
 EL+SRKVKFLLDVPQQDGGQWDMVV+LFEKYGVVPK+VYPES+SSS SRELNQVLNKLRL
 Sbjct: 121 ELTSRKVKFLLDVPQQDGGQWDMVSLFEKYGVVPKSVYPESISSNSRELNQVLNKLRL 180

Query: 181 QDAQILRELIAQAGDGVQNKKEELLQEIFNFLAMNLGLPPQSFDFAVRDKDNHYQSDK 240
 QDAQILR+LIA GA V+++K ELLQEIFNFLAM LGLPP+ FDFAYRDKD+HY +K
 Sbjct: 181 QDAQILRLIASGAKADQVEDRKABELLQEIFNFLAMTLGLPPRHDFAYRDKDDHYHVEK 240

Query: 241 NITPKAFYQKYVNLDSLSDYVSIINAPTVDKPYGQSYTVEMLGNVVGPAVKYLNLDMKRF 300
 +TP+AFY K+V L LSDYVS+INAPT DKPYG+SYTVEMLGNVVG V+YLNLDMKRF
 Sbjct: 241 GLTPQAFYDKFVGLKLSLDYVSVINAPTADKPYGKSYTVEMLGNVVGSRVRYLNLDMKRF 300

Query: 301 KELAIQMKSGETVWFGSDVGQVSNRQKGILATTTYDFNSSMDIKLSQDKAGRLDYSESL 360
 KELAI QM++GE+VWFGSDVGQVS+RQKGILAT TYDF +SMDI LSQDKAGRLDYSESL
 Sbjct: 301 KELAIKQMAGESVWFGSDVGQVSDRQKGILATNTYDFEASMDINLSQDKAGRLDYSESL 360

Query: 361 MTHAMVLTGVLDLDESGQPLKWKVENSWEKVGDKGYFVASDAWMDEYTYQIVVRKELLTK 420
 MTHAMVLTGVLDLDE+G+PLKWKVENSWEKVG GYFVASDAWMDEYTYQIVVRKE LT
 Sbjct: 361 MTHAMVLTGVLDLDETGKPLKWKVENSWEKVGDKGYFVASDAWMDEYTYQIVVRKEFLTA 420

Query: 421 EELEAYNAEPITLAPWDPMGALA 443
 +EL AY EP LAPWDPMGALA
 Sbjct: 421 DELAAYEKEPQVLAPWDPMGALA 443

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 965

A DNA sequence (GBSx1024) was identified in *S.agalactiae* <SEQ ID 2947> which encodes the amino acid sequence <SEQ ID 2948>. Analysis of this protein sequence reveals the following:

Possible site: 36
 >>> Seems to have a cleavable N-term signal seq.

-1061-

----- Final Results -----

bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

5

A related GBS nucleic acid sequence <SEQ ID 9533> which encodes amino acid sequence <SEQ ID 9534> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

10 >GP:AAF17262 GB:AF210752 penicillin-binding protein 1A
 [Streptococcus pneumoniae]
 Identities = 412/725 (56%), Positives = 544/725 (74%), Gaps = 14/725 (1%)

15 Query: 4 IKKESVIKLLKYAFGIIMGFIILAIIVIGLLFAYYVSRSPKLTQALKSVNSSLVYDGN 63
 + K +++L+KY + +I AIV+GG +F YYVS++P L++ L + SS +YD N
 Sbjet: 1 MNKPITLRLIKYLSISFLSLVIAIIVLGGGVFFYYVSKAPSLSESKLVAITSSKIYDNKN 60

20 Query: 64 KLIADLGSEKRESVSADSIPLNLVNAITSIEDKRFFKHRGVDIYRILGAAWHNLVSSNTQ 123
 +LIADLGSE+R + A+ IP +LV AI SIED RFF HRG+D RILGA NL S++ Q
 Sbjet: 61 QLIADLGSERRVNAQANDIPTDLVKAIVSIEDHRFFDHRGIDTIRILGAFLRNLQSNLSQ 120

25 Query: 124 GGSTLDQQLIKLAYFSTNKSQDTLKRKSQEVWLALQMERKYTKEEILTFYINKVYMGNGN 183
 GGSTL QQLIKL YFST+ SDQT+ RK+QE WLA+Q+E+K TK+EILT+YINKVYM NGN
 Sbjet: 121 GGSTLTQQLIKLTYFSTSTSDQTISRKAQEAWLAIQLEQKATKQEILTYYINKVYMSNGN 180

30 Query: 184 YGMRTTAKSYFGKDLKELSLIAQLALLAGIPQAPTQYDPYKNPESAQTRRNTVLQOMYQDK 243
 YGM+T A++Y+GKDL LS+ QLALLAG+PCAP QYDPY +PE+AQ RRN VL +M
 Sbjet: 181 YGMQTAAQNYGKDLNLSLPQLALLAGMPQAPNQYDPYSHPEAAQDRRNVLSEMKNQG 240

35 Query: 244 NISKKEYDQAVATPVTDGLKELKQKSTYPKYMDNYLKQVISEVKQKTGKDIFTAGLKVYT 303
 IS ++Y++AV TP+TDGL+ LK S YP YMDNYLK+VI++V++TG ++ T G+ VYT
 Sbjet: 241 YISAEQYKAVNTPITDGLQSLKSASNPAYMDNYLKEVINQVEEETGYNLLTTGMDVYT 300

40 Query: 304 NINTDAQQLYDIYNSDTYIAYPNNELQIASTIMDATNGKVIAQLGGRHQENISFGTNQ 363
 N++ +AQK L+DIYN+D Y+AYP++ELQ+ASTI+D +NGKVIAQLG RHQ+ N+SFG NQ
 Sbjet: 301 NVDQEAQKHLWDIYNTDEYVAYPDDELQVASTIVDVSNQKVIAQLGARHQSSNVSFQINQ 360

45 Query: 364 SVLTD RDWGSTMKPI SAYAPAI DSGVYNSTGQSLNDSVYYWPGTSTQLYDWD RQYMGWMS 423
 +V T+RDWGSTMKPI+ YAPA++ GVIY+ST ++D Y +PGT T +Y+WDR Y G ++
 Sbjet: 361 AVETNRD WGSTMKPI TDYAPALEYGVYDSTATIVHDEPYNYPGTIDTPVYNWDRGYFGNIT 420

50 Query: 424 MQTAIQQSRNVPVRALEAAGLDEAKSFLEKLGIIYPENYNSNAISSNNSSSDAKYGASS 483
 +Q A+QSRNVPV L GL+ AK+FL LGI YP ++YSNAISSN + SD KYGASS
 Sbjet: 421 LQYALQQSRNVPVETLNVGLNRAKTFNLGLGIDYPSLHYNSNAISSNTTESDKKYGASS 480

55 Query: 484 EKMAAAYSAFANGGTYYPQYVKNKIEFSDGTNDTYAASGSRAMKETTAYMMTDMKTTLV 543
 EKMAAAY+AFANGGTYYP Y++K+ FSDG+ ++ G+RAMKETTAYMMTDM+KTTL
 Sbjet: 481 EKMAAAYAFANGGTYYPKPMYIHKVVFSDGSEKEFSNVGTRAMKETTAYMMTDMKTTLV 540

60 Query: 544 FGTGTPKAAIPGVAQAGKTGTSTNYTEDELAKEATGTGIYNSAVGTMAPDENFVGYSKYTM 603
 +G G A +P + QAGKTGTSTNYT++E+ K Y G +APDE FVGYT KY M
 Sbjet: 541 YGIGRGAYLPWLPQAGKTGTSTNYTDEEIEK-----YIKNTGYVAPDEMFGYTRKYAM 593

Query: 604 AIWTGYKNRLTPLYGSQLDIATEVYRAMSYLTGGYSA-DWTMPEGLYRSGSYLYINGTT 662
 A+WTGY NRLTPL G L +A +VYR+MM+YL+ G + DW +PEGLYR+G +++ NG
 Sbjet: 594 AVWTGYSNRLTPLVGDGLTVAKVYRSMYTLSEGSNPEDWNIPGLYRNGEFVFKNGAR 653

Query: 663 TTGTYSVVYKNIYQNSGQSSQSSSSSTSEKQKEDKNTANDANSSSPQVETPNNGNATTP 722
 +T +SS + S +SS SSS +S+ + + N++ +++P T + TTP
 Sbjet: 654 ST--WSSPAPQQ--PPSTESSSSSSSSSTSQSNSTTPSTNNSTTTNPNNTQQSN--TTP 707

Query: 723 NNSNQ 727
 + NQ
 Sbjet: 708 DQQNQ 712

-1062-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2949> which encodes the amino acid sequence <SEQ ID 2950>. Analysis of this protein sequence reveals the following:

```

Possible site: 41
>>> Seems to have an uncleavable N-term signal seq
5   INTEGRAL    Likelihood =-13.96    Transmembrane    19 - 35 ( 9 - 43)

----- Final Results -----
          bacterial membrane --- Certainty=0.6583(Affirmative) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
10         bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

>GP:CAA88918 GB:Z49095 penicillin-binding protein 1a [Streptococcus pneumoniae]
Identities = 422/712 (59%), Positives = 536/712 (75%), Gaps = 8/712 (1%)
15
Query: 4   IKNPKILKWLKYLVSAILSLIILVIIIGGLLFTFYISSAPKLSEAQLKSTNSSLVYDGNN 63
      + P IL+ +KY+ + LSL+I I++GG +F +Y+S AP LSE++L +T SS +YD N
Sbjct: 1   MNKPTILRLIKYLSISFLSLVIAAIVLGGGVFFYYVSKAPSLSESKLVAATTSSKIYDNKN 60

20
Query: 64   NLIADLGSEKRENTADSIPINLVNAITSIEDKRFFNHRGVDLYRIFGAFFHNLTSTQTQ 123
      LIADLGSE+R N A+ IP +LV AI SIED RFF+HRG+D RI GA NL S + Q
Sbjct: 61   QLIADLGSERRVNAQANDIPTDLVKAIVSIEDHRFPDHRGIDTIRILGAFLRNLSNSLIQ 120

25
Query: 124  GGSTLDQQLIKLAYFSTNESDQTLKRKAQEVWLALQMERKYTKQEILTTFYINKVYMGNGN 183
      GGSTL QQLIKL YFST+ SDQT+ RKAQE WLA+Q+E+K TKQEILT+YINKVYM NGN
Sbjct: 121  GGSTLTQQLIKLYFSTSTSDQTIISRKAQEAWLAIQLEQKATKQEILTYINKVYMSNGN 180

30
Query: 184  YGMLTAAKSYYGKDLKDLSYAQLALLAGIPQAPSQYDPYLHPAAQNRNRNVVLQOMYMEK 243
      YGM TAA++YYGKDL +LS QLALLAG+PQAP+QYDPY HPEAAQ+RRN+VL +M +
Sbjct: 181  YGMQTAAQNYYGKDLNLSLPQLALLAGMPQAPNQYDPYSHPEAAQDRRNLVLSEMKNQG 240

35
Query: 244  HLTKAHEYETAATPVAEGLQSLQQRSTYPKYMDNYLKQVIEEVKKETNKDIFTAGLKVYT 303
      +++ +YE A+ TP+ +GLQSL+ S YP YMDNYLK+VI +V++ET ++ T G+ VYT
Sbjct: 241  YISARQYEKAVNTPITDGLQSLKASNPAYMDNYLKEVINQVEEETGYNLLTGTGMDVYT 300

40
Query: 304  NIIPDAQQTLYNIYHSGDYVYYPDQDFQVASTIVDVTNGHVIAQLGGRNQDENVSFGTNQ 363
      N+ +AQ+ L++IY+S YV YPD D QVAST+VDV+NG VIAQLG R+Q NVSFGTNQ
Sbjct: 301  NVDQEAQKHLWDIYNSDQYVSYPDDDLQVASTVVDVSNQKQVIAQLGARHQASNVSGTNQ 360

45
Query: 364  AVLTDRDWGSTMKPITAYAPAIESGVYTSTAQSTINDSVYYWPGTTTTQLFNWDLRYNGWMT 423
      AV T+RDWGS+MKPIT YAPA+E GYV STA +D Y +PGT T L+NWD Y G +T
Sbjct: 361  AVETNRDWGSSMKPITDYAPALEYGVYDSTASIVHDVPYNYPGTDTPLYNWDHVYFGNIT 420

50
Query: 424  IQAAIMLSRNVPAVRALEAAGLDYARSFLSSLGINYPEMHYSNAISSNNSSSDKKYGASS 483
      IQ A+ SRNV AV L GLD A++FL+ LGI+YP MHY+NAISSN + S+KKYGASS
Sbjct: 421  IQYALQQSRNVTAVEFLNKVGLDRAKTFNLGLGIDYPSMHYANAISSNTTESNKKYGASS 480

55
Query: 484  EKMAAAYAAFANGGIYHKPRYVNKVEFSDGTSKTFDEKGRAMKETTAYMMTDLKTIVLT 543
      EKMAAAYAAFANGGIYHKP Y+NK+ FSDG+ K F + G RAMKETTAYMMT+M+KTIVLT
Sbjct: 481  EKMAAAYAAFANGGIYHKPMYINKIVFSDGSEKEFSDAGTRAMKETTAYMMTEMMKTIVLT 540

60
Query: 544  YGTGTAAAI PGVAQAGKTGTSNYTDEELAKIGEKYGLYPDYVGTILAPDENFVGFTKRYAM 603
      YGTG A +P + QAGKTGTSNYTDEE+ K Y G +APDE FVG+T++YAM
Sbjct: 541  YGTGRGAYLPWLPQAGKTGTSNYTDEEIEK-----YIKNTGYVAPDEMFGVYTRKYAM 593

65
Query: 604  AVWTGYKNRLTPVYGSSLEIASDVYRSMPTYLT-NGYSEDWTMPNGLYRSGGFLYLSGTI 662
      AVWTGY NRILT+ G +A VYRSM+TYL+ + DWTMP+GLYR+G F++ +G
Sbjct: 594  AVWTGYSNRILTPIIGDGLVAGKVYRSMITYLSEDDQPGDWTMPDGLYRNGEFVFKNGAR 653

70
Query: 663  ASNTDYTNVYNLYSNNTTTASSQTSDDTSSSNDTSNSTNTDNNNGSHPS 714
      ++ + + S++++ SS + S+ T+ S + S +TN +NN +T
Sbjct: 654  STWSSPAFQPPSTESSSSSSDSSTSQSNSSTTPSTNNSTTNPNNNTQQSNT 705

```

An alignment of the GAS and GBS proteins is shown below.

-1063-

Identities = 521/729 (71%), Positives = 621/729 (84%), Gaps = 10/729 (1%)

Query: 1 MITIKKESVIKLLKYAFGIIMGFIIILAIIVIGLLLFAYYVSRSPKLTDOALKSVNSSLVYD 60
 +ITIK ++K LKY I+ IIL I+IGGLLF +Y+S +PKL++ LKS NSSLVYD
 5 Sbjet: 1 VITIKNPKILKWLKYVLSAILSLIILVIIIGGLLFTFYISSAPKLSEAQLKSTNSSLVYD 60

Query: 61 GNNKLIADLGSEKRESVSADSIPLNLVNAITSIEDKRFFKHGVDIYRILGAAWHNLVSS 120
 GNN LIADLGSEKRE+V+ADSIP+NLVNAITSIEDKRFF HRGVD+YRI GAA+HNL S
 10 Sbjet: 61 GNNNLIADLGSEKRENTADSIPINLVNAITSIEDKRFFNHRGVDLYRIFGAAAFHNLTSQ 120

Query: 121 NTQGGSTLDQQLIKLAYFSTNKSDQTLKRKSQEVWLALQMERKYTKKEILTFFYINKVYMG 180
 TQGGSTLDQQLIKLAYFSTN+SDQTLKRK+QEVWLALQMERKYTK+EILTFFYINKVYMG
 15 Sbjet: 121 TTQGGSTLDQQLIKLAYFSTNESDQTLKRKAQEVWLALQMERKYTKQEILTFFYINKVYMG 180

Query: 181 NGNYGMRTTAKSYFGKDLKELSLIAQLALLAGIPQAPTQYDPYKNPESAQTRRNTVLQQMY 240
 NGNYGM T AKSY+GKDLK+LS AQLALLAGIPQAP+QYDPY +PE+AQ RRN VLQQMY
 20 Sbjet: 181 NGNYGMLTAAKSYYGKDLKDLSYAQLALLAGIPQAPSQYDPYLHPEAAQNRNRNVVLQQMY 240

Query: 241 QDKNISKKEYDQAVATPVTDLGKELKQKSTYPKYMDNYLKQVISEVKQKTGKDIFTAGLK 300
 +K+++K EY+ A+ATPV +GI+ L+Q+STYPKYMDNYLKQVI EVK++T KDIFTAGLK
 25 Sbjet: 241 MEKHLTKAEYETAATPVAEGLQSLQQRSTYPKYMDNYLKQVIEEVKKETNKDIFTAGLK 300

Query: 301 VYTNINTDAQQLYDIYNSDTYIAYPNNELQIASTIMDATNGKVIAQLGGRHQENISFG 360
 VYTNI DAQ+ LY+IY+S Y+ YP+ + Q+ASTI+D TNG VIAQLGGR+Q+EN+SFG
 30 Sbjet: 301 VYTNIIIPDAQQTLYNIYHSGDYVYYPDQDFQVASTIVDVTNHGVIAQLGGRNQDENVSFG 360

Query: 361 TNQSVLTDRDWGSTMKPISAYAPAIIDSGVYNSTGQSLNDSVYYWPGTSTQLYDWRQYMG 420
 TNQ+VLTDRDWGSTMKPI+AYAPAI+SGVY ST QS NDSVYYWPGT+TQL++WD +Y G
 35 Sbjet: 361 TNQAVLTDRDWGSTMKPIAYAPAIESGVYTSTAQSTNDSVYYWPGTTTQLFNWDLRYNG 420

Query: 421 WMSMQTAIQSRNVPVRALEAAGLDEAKSFLEKLGIIYYPEMNYSNAISSNNSSSDAKYG 480
 WM++Q AI SRNVPVRALEAAGLD A+SFL LGI YPEM+YSNAISSNNSSSD KYG
 40 Sbjet: 421 WMTIQAAIMLSRNVPVRALEAAGLDYARSFLLSLGINYPEMHYSNAISSNNSSSDKKYG 480

Query: 481 ASSEKMAAAYSAFANGGTYYKPYVKNKIEFSDGINDTYAASGSRAMKETTAYMMTDLTKT 540
 ASSEKMAAAY+AFANGG Y+KP+YVKN+EFSDGT+ T+ G RAMKETTAYMMTDLTKT
 45 Sbjet: 481 ASSEKMAAAYAFANGGIYHKPRYVKNKVEFSDGTSKTFDEKGRAMKETTAYMMTDLTKT 540

Query: 541 VLTFTGTGTAAPGVAQAGKTGTSNYTEDELAKEATGTGIYNSAVGTMAPDENFVGYSK 600
 VLT+GTGT AAIPGVAQAGKTGTSNYT++ELAKI G+Y VGT+APDENFVG+T +
 50 Sbjet: 541 VLTFTGTGTAAPGVAQAGKTGTSNYTDEELAKIGEKYGLYPDYVGTLPDENFVGFTKR 600

Query: 601 YTMAIWTGYKNRLTPLYGSOLDIATEVYRAMMSYLTGGYSADWTIMPEGLYRSGSYLYING 660
 Y MA+WTGYKNRLTP+YGS L+IA++VYR+MM+YLT GYS DWTMP GLYRSG +LY++G
 55 Sbjet: 601 YAMAVWTGYKNRLTPVYGSSLEIASDVYRSMMTYLTNGYSEDWTMPNGLYRSGGFLYLSG 660

Query: 661 TTTTGT-YSSSVYKNYQNSGQSSQSSSSTSSEKQKEDKNTANDANSSSPQVETPNNGNA 719
 T + T Y++SVY N+Y N ++++ SS+ +D +++ND ++S+ T NNG+
 60 Sbjet: 661 TYASNTDYTNVYNLYSN-----NTTTASSQTSDDTSSSNDTSNST---NTDNNGSH 711

Query: 720 TTPNNSNQT 728
 + ++ T
 65 Sbjet: 712 PSTDDKKT 720

55 A related GBS gene <SEQ ID 8695> and protein <SEQ ID 8696> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 10
 McG: Discrim Score: 6.55
 60 GvH: Signal Score (-7.5): -1.98
 Possible site: 36
 >>> Seems to have a cleavable N-term signal seq.
 ALOM program count: 0 value: 4.03 threshold: 0.0
 PERIPHERAL Likelihood = 4.03 201
 modified ALOM score: -1.31
 65 *** Reasoning Step: 3

-1065-

```

2157      2187      2214      2244      2274      2304      2334      2364
KNRLTPLYGSQLDIATEVYRAMMSYLT-GGYSADWTMPEGLYRSGSYLYINGTTTTTGTYSVVYKNIYQNSGSSQSSSS
||||| | | :| :|||:|:|:| | | :|||:|:|:| | :| :| | :| | | |
5      SNRLTPLVGDGLTVAAKVYRSMMTYLSSESNPEDWNIPEGLYRNGEFVFKNGARST--WSSPAPQQ--PPSTESSSSSSD
      610      620      630      640      650      660      670

2394      2424      2454      2484      2514      2544      2574      2604
TSSEKQKEDKNTANDANSSSPQVETPNNGNAITPNNSNQTVPGTGHGNGNGNNTVPNGN*KTGYIIQFFNL*LLLLLI
:|: : : | :|: :| | : | | | : | |
10     SSTSQSNSTTPSTNNSTTTNPNNTQQS--NITPDQQNQNPQPAQP
      690      700      710

```

SEQ ID 8696 (GBS146) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 23 (lane 4; MW 82kDa), in Figure 168 (lane 11-13; MW 96.5kDa) and in Figure 238 (lane 8; MW 96.5kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 49 (lane 2; MW 107kDa).

Purified Thio-GBS146-His is shown in Figure 244, lane 4.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 966

A DNA sequence (GBSx1025) was identified in *S.agalactiae* <SEQ ID 2951> which encodes the amino acid sequence <SEQ ID 2952>. Analysis of this protein sequence reveals the following:

```

Possible site: 37
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3647(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:AAA26957 GB:M90528 ORF [Streptococcus oralis]
Identities = 143/196 (72%), Positives = 165/196 (83%), Gaps = 1/196 (0%)

Query: 1  MVNYPHQLIRKTTVTKSKKKKIDFANRGMSFEAAINATNDYYLSHELAVIHKKPTPVQIV 60
      +  +  +  +  +  +  +  +  +  +  +  +  +  +  +  +  +  +  +  +  +  +  +  +
Sbjct: 1  MVNYPHKISSQKRQAPPSQTK-NFANRGMSFEKMINATNDYYLSHGLAVIHKKPTPIQIV 59

Query: 61  KVDYPKRSRAKIVEAYFRQASTTDYSGVYKGYIDFEAKETRQKTAMPKMFHQAIEHM 120
      +VDYP+RSRAKIVEAYFRQASTTDYSGVY GYYIDFEAKETRQK A+PMKNFH HQI+HM
Sbjct: 60  RVDYPQRSRAKIVEAYFRQASTTDYSGVYDGYIDFEAKETRQKHAIPMKNFHHHQI QHM 119

Query: 121 ANVLQQKGICFVLLHFSTLKETYLPAANELISFYQIDKGNKSMPIYIRKNGFFVKESAF 180
      VL Q+GICFVLLHF++ +ETYLPA +LI FY DKG KSMP+ YIR+NG+ ++ AF
Sbjct: 120 EQVLAQRGICFVLLHFASQOETYLPAVDLIRFYHQDKGQKSMPLGYIRENGYRIELGAF 179

Query: 181 PQVPYLDIIEKLLGG 196
      PQ+PYLDII+E LLGG
Sbjct: 180 PQIPYLDIIEKLLGG 195

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2953> which encodes the amino acid sequence <SEQ ID 2954>. Analysis of this protein sequence reveals the following:

```

Possible site: 37
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

```

-1066-

bacterial cytoplasm --- Certainty=0.5030(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

5 An alignment of the GAS and GBS proteins is shown below.

Identities = 166/199 (83%), Positives = 177/199 (88%)

Query: 1 MVNYPHQLIRKTTVTKSKKKKIDFANRGMSFEAAINATNDYYLSHELAVIHKKTPVQIV 60
 MVNYPH LIR+ + K+ K+DFANRGMSFEAAINATNDYYLS ++AVIHKKTPVQIV
 10 Sbjct: 1 MVNYPHNLIRQKVSSVQKQNVDFANRGMSFEAAINATNDYYLSRQIAVIHKKTPVQIV 60

Query: 61 KVDYPKRSRAKIVEAYFRQASTTDYSGVYKGYIIDFEAKETRQKTAMPKKNFHAHQIEHM 120
 KVDYPKRSRAKIVEAYFRQASTTDY GVIKG+Y+DFEAKETRQKTAMPKKNFH HQIEHM
 15 Sbjct: 61 KVDYPKRSRAKIVEAYFRQASTTDYCGVYKGYHYVDFAKETRQKTAMPKKNFHLHQIEHM 120

Query: 121 ANVLQKKGICFVLLHFSTLKETYLIPANELISFYQIDKGNKSMPIIDYIRKNGFFVKESAF 180
 A VL QKGICFVLLHFSTLKETY LPA LISFYQID G+KSMPIIDYIRKNGF V AF
 Sbjct: 121 ACVLHQKKGICFVLLHFSTLKETYYLPAQALISFYQIDNGSKSMPIIDYIRKNGFKVAFGAF 180

20 Query: 181 PQVPYLDIIEBKLLGGDYN 199
 PQVPYL+IIE+ LGGDYN
 Sbjct: 181 PQVPYLNIEQNFLGGDYN 199

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 25 vaccines or diagnostics.

Example 967

A DNA sequence (GBSx1026) was identified in *Sagalactiae* <SEQ ID 2955> which encodes the amino acid sequence <SEQ ID 2956>. Analysis of this protein sequence reveals the following:

Possible site: 61
 30 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3227(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 35 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB14136 GB:Z99115 similar to hypothetical proteins from B. subtilis [Bacillus subtilis]
 40 Identities = 74/174 (42%), Positives = 97/174 (55%), Gaps = 6/174 (3%)

Query: 5 ILVTGYKNFELGIFQDKDPRITIIKKAIDKDFRFLENGADWFIFMGNLGFYWALEVAL 64
 + +TGYK FELGIF+ D + IKKAI FL+ G +W + G LG E WA E A
 45 Sbjct: 4 LAITGYKPFELGIFQDDKALYYIKKAIKNRLIAFLDEGLEWILISGQLGVELWAAEAAAY 63

Query: 65 DLQKEY-DFQIATIFTFENHGQNWNEANKAKL-ALFKQVDF-VKYTFPSYENPGQFKQYN 121
 DLQ+EY D ++A I F +NW E NK + A+ Q D+ T YE+P QFKQ N
 Sbjct: 64 DLQEEYFDLKVAVITPFYEQBNWKEPNKEQYEAFLAQADYEASLTHRPYESPLQFKQKN 123

50 Query: 122 HFLINNTQGAYLFYDSENETNLKFLLEMEKK---EAYDISFLTFRDLNEIYEE 172
 F I+ + G L YD E E + K++L EK+ + Y I F+T D L EE
 Sbjct: 124 QFFIDKSDGLLLLYDPEKEGSPKMYLGTAEKRREQDGYPIYFITMDDLRTVTEE 177

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2957> which encodes the amino acid
 55 sequence <SEQ ID 2958>. Analysis of this protein sequence reveals the following:

Possible site: 60
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

-1067-

bacterial cytoplasm --- Certainty=0.3041(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

5 An alignment of the GAS and GBS proteins is shown below.

Identities = 102/167 (61%), Positives = 127/167 (75%)

Query: 3 STILVTGYKNFELGIFQDKDPRITIIKKAIDKDFRRFLENGADWFIFMGNLGFEYWALEV 62
 + IL+TGY++FE+GIF KDPR++IIK+AI KD +LENG DWFIF GNLGFE WALEV
 10 Sbjct: 2 TAILITGYRSFEIGIFDHDKDFRVSIIKQAIRKDLIGYLENGVDWFIFTGNLGFEQWALEV 61

Query: 63 ALDLQKEYDFQIATIFTFENHCQNWNEANKAKLALFKQVDFVKYTFPSYENPGQFKQYNH 122
 A +L+++EY QIATIF FE HG WNE NK L+ P+ VDFVKY FP+YE P QF QY
 15 Sbjct: 62 ANELKEEYPLQIATIFLFEETHGDRWNEKNKEVLSQFRAVDFVKYFPNYEQPTQFSQYYQ 121

Query: 123 FLINNTQGAYLFYDSENETNLKFLLEMEKKEAYDISFLTDFRLNEI 169
 FL+ T+GAY+FYD+ENETNLK+ L+ + Y+ LTFDRLN++
 Sbjct: 122 FLLEKTEGAYVFYDTENETNLKYFLKKAIDMPHYQLLLLTDFRLNDM 168

20 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 968

A DNA sequence (GBSx1027) was identified in *S.agalactiae* <SEQ ID 2959> which encodes the amino acid sequence <SEQ ID 2960>. Analysis of this protein sequence reveals the following:

25 Possible site: 23
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 30 bacterial cytoplasm --- Certainty=0.5188(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

35 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 969

A DNA sequence (GBSx1028) was identified in *S.agalactiae* <SEQ ID 2961> which encodes the amino acid sequence <SEQ ID 2962>. This protein is predicted to be cell division protein DivIVA. Analysis of this
 40 protein sequence reveals the following:

Possible site: 16
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 45 bacterial cytoplasm --- Certainty=0.2736(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9535> which encodes amino acid sequence <SEQ ID 9536>
 50 was also identified.

The protein has homology with the following sequences in the GENPEPT database.

-1068-

>GP:CAB14135 GB:Z99115 ypsB [Bacillus subtilis]
 Identities = 46/102 (45%), Positives = 69/102 (67%), Gaps = 14/102 (13%)

Query: 14 SPKDIFEQDFKVS MRGYDKKEVDVFLDDVIKDYENYLEQIEKLQ MENRRLQQALDKKKESE 73
 S K+I E++FK +RGY +++VD FLD +IKDYE + ++IE+LQ EN +L++ L+ E

Sbjct: 9 SAKEILEKEFKTGVRGYKQEDVDKFLDMIKDYETFHQEI EELQQENLQLKKQLE----E 64

Query: 74 ASNVNRSGTAMYNQKPIAQSATNFDILKRISLEKEVFG RQI 115
 AS ++P+ + TNFDILKR+S LEK VFG ++

Sbjct: 65 AS-----KKQPVQSNITNFDILKRISLEKEHVF GSKL 96

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2963> which encodes the amino acid sequence <SEQ ID 2964>. Analysis of this protein sequence reveals the following:

Possible site: 51
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4466(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 71/112 (63%), Positives = 85/112 (75%), Gaps = 6/112 (5%)

Query: 8 MASIIYSPKDIFEQDFKVS MRGYDKKEVDVFLDDVIKDYENYLEQIEKLQ MENRRLQQAL 67
 M SIIYSPKDIFEQ+FK SMRG+DKKEVD FLD+VIKDYEN+ QIE L+ EN +AL

Sbjct: 1 MTSIIYSPKDIFEQEFKTS MRGFDKKEVDFLDNVIKDYENFNAQIEALKAEN----EAL 56

Query: 68 DKKESEASNVNRSGTAMYNQKP--IAQSATNFDILKRISLEKEVFG RQIRE 117
 K + +A N ++ +P +AQSATNFDILKRIS+LEKEVFG+QI E

Sbjct: 57 KKAKFQARNTVSATVQQPVQPTRVAQSATNFDILKRISLEKEVFGKQIIE 108

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 970

A DNA sequence (GBSx1029) was identified in *S.agalactiae* <SEQ ID 2965> which encodes the amino acid sequence <SEQ ID 2966>. Analysis of this protein sequence reveals the following:

Possible site: 55
 >>> Seems to have no N-terminal signal sequence (or aa 1-19)

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0655(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB14134 GB:Z99115 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 204/382 (53%), Positives = 274/382 (71%), Gaps = 3/382 (0%)

Query: 3 ESFKLIATAAAGLEAIVGREIRNLGIDCQVENGRVRFHGDIKTIIETNLWLR AADRIKII 62
 + + LIATA G+EA+V +E+R+LG +C+V+NG+V F GD I NLWLR ADRIK+

Sbjct: 2 KKYTLIATAPMGIEAVVAKEVRDLGYECKVDNGKVI FEGDALAICRANLWLR TADRIKQV 61

Query: 63 VGEFPAPTFEELFQGVYGLDWENYLP LGAKFPIAKAKCVKSKLHNEPSVQAISKKAVAKK 122
 V F A TF+ELF+ ++W +++P KFP+ K VKS L + P Q I KKA+ +K

Sbjct: 62 VASFKAKTFDELFEKTKAINWRSFIPENGKFPVI-GKSVKSTLASVPDCQRIVKKAIVEK 120

Query: 123 LQKVFHRPEGVPLQENGAEFKIEVSILKDKATVMIDTTGSSSLFKRGYRAEKG GAPIKENM 182
 L K+ ++E GAE+K+E+S+LKD+A + +D++G+ L KRGYR ++GGAPIKE +

-1069-

Sbjct: 121 L-KLQSGKANDWIBETGAEYKVEISLLKDQALITLDSSTGLHKG YRVDQGGAPIKETL 179

Query: 183 AAATIQLSNWFDPKPLIDPTCGSGTFCIEAAMIGMNIAPGFNRDFAFEAWPWVDSQVQK 242
 AAA++QL+NW PD+P +DP CGSGT IEAA+IG NIAPGFNRDF E W W+ + K

5 Sbjct: 180 AAALVQLTNWTPDRPFVDFCGSGTIAIEAALIGQNIAPGFNRDFVSEWVWIGKDLWNK 239

Query: 243 VRDEAESKANYDIDLDISGFDLDGRMVEIARKNAEEAGLDVILKQMRQLQDLKTDKING 302
 R E E KANYD L I D+D RMV+IA++NAEEAGLD+I+ KQM+++D T+ G

10 Sbjct: 240 ARLEVEEKANYDQPLTTFASDIDHRMVQIAKENAEEAGLDLIQFKQMVKDFTTNLEFG 299

Query: 303 VIISNPPYGERLLDDKAVIDILYNEMGQTFAPLKTWSKFILTSDEGFEKKYGSQADKKRKL 362
 VI+ NPPYGERL + KAV+ +Y EMGQ F PL TWS ++LTS+E FE+ YG +A KKRKL

Sbjct: 300 VIVGNPPYGERLGEKKAVEQMYKEMGQAFELDTWSVYMLTSNENFEEAYGRKATKKRKL 359

15 Query: 363 YNGTLKVDLYQYYGERVRRQVK 384
 +NG +K D YQY+ +VR Q K

Sbjct: 360 FNGFIKTDYQYW-SKVRPQRK 380

20 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2967> which encodes the amino acid
 sequence <SEQ ID 2968>. Analysis of this protein sequence reveals the following:

Possible site: 14
 >>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0324 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

30 Identities = 317/383 (82%), Positives = 354/383 (91%)

Query: 1 MKESFKLIATAAGLEAIVGREIRNLGIDCQVENGRVRFHGDIKTIIETNLWLRAADRIK 60
 MKE+F+L+ATAAGLEA+VG+E+R LG DCQVENG+V F GD++ I++TNLWLRAADRIK

35 Sbjct: 1 MKETFRLVATAAGLEAVVGKEVRALGFDCQVENGVYFEGDVEAIVKTNLWLRAADRIK 60

Query: 61 IIVGEFPAPTFEELFQGVYGLDWNVPLGAKFPIAKAKCVKSKLHNEPSVQAIKKAVA 120
 IIVG+FP A TFEELFQGV+ LDWENVPLGAKFPI+KAKCVKSKLHNEPSVQAI+KKAV

Sbjct: 61 IIVGQFPAPTFEELFQGVFALDWNVPLGAKFPIKAKCVKSKLHNEPSVQAITKKAVV 120

40 Query: 121 KKLQKVFHRPEGVPLQENGAEFKIEVSILKDKATVMIDTTGSSSLFKRGYRAEKGAPIKE 180
 KKLQK FHRPEGVPLQE G+ F IEVSILKD+AT+MIDTTGSSSLFKRGYR +KKGAPIKE

Sbjct: 121 KKLQKHFRPEGVPLQEVGSTFNIEVSILKDQATIMIDTTGSSSLFKRGYRVQKKGAPIKE 180

45 Query: 181 NMAAAIIQLSNWFDPKPLIDPTCGSGTFCIEAAMIGMNIAPGFNRDFAFEAWPWVDSQV 240
 NMAAAI+ LSNWFDPKPL+DPTCGSGTFCIEAAMIGMNIAPGFNR FAFE W WVD+ V

Sbjct: 181 NMAAAILALSNWFDPKPLVDPTCGSGTFCIEAAMIGMNIAPGFNRSF AFEWVWVDKDMV 240

50 Query: 241 QKVRDEAESKANYDIDLDISGFDLDGRMVEIARKNAEEAGLDVILKQMRQLQDLKTDKI 300
 Q+VRD+AE +ANY+I+LDISGFD+DGRM+EIA+ NAEEAGL DVI KQMRQLD +TDK+

Sbjct: 241 QQVRDDAEQEANYEIELDISGFDIDGRMIEIAKSNAEEAGLSDVITFKQMRQLQDFRTDKV 300

55 Query: 301 NGVIISNPPYGERLLDDKAVIDILYNEMGQTFAPLKTWSKFILTSDEGFEKKYGSQADKKR 360
 NGV+ISNPPYGERLLDDKAVIDILYNEMGQTFAPLKTWSKFILTSDE FE KYG +ADKKR

Sbjct: 301 NGVVISNPPYGERLLDDKAVIDILYNEMGQTFAPLKTWSKFILTSDELFEKYGQKADKKR 360

Query: 361 KLYNGTLKVDLYQYYGERVRRQV 383
 KLYNGTLKVDLYQ+YGERV+R +

Sbjct: 361 KLYNGTLKVDLYQFYGERVKRHL 383

60 SEQ ID 2966 (GBS255) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell
 extract is shown in Figure 43 (lane 7; MW 44kDa). It was also expressed in *E.coli* as a GST-fusion
 product. SDS-PAGE analysis of total cell extract is shown in Figure 48 (lane 4; MW 69kDa).

-1071-

Sbjct: 229 ---DSSDSSNSSESSDSSD--SSDSDSSDSSDSSNSNSDSDS--SNSDSDSSNSDSSD 283

Query: 298 AVVDGKKSEKLEVKDGANFDSLDSKTLNTGNASLDSLHHSIVSTGRNQVKQSEEQASSNK 357
+ S+ + D +N S DS + + S DS S + N S+ SS+

Sbjct: 284 SSDSSNSDSSDSSDSSN--SSDSDSSDSSDSSNSDSDSSNSDSSDSSDSSDSSDSSD 341

Query: 358 VSDTQITEQPNVTNGQSSSSAATINNQAAGTASGNLERNRS 398

+ + ++ + ++ SS+S+ + N+ + + + + S

Sbjct: 342 SNSSDSSDSSDSSDSSNSDSSNSDSSDSSDSSDSSDSSDSSDSSDSSDSSDSSD 382

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2971> which encodes the amino acid sequence <SEQ ID 2972>. Analysis of this protein sequence reveals the following:

Possible site: 28

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -14.70 Transmembrane 180 - 196 (175 - 202)

----- Final Results -----

bacterial membrane --- Certainty=0.6880(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAF15293 GB:AF202180 erythrocyte membrane-associated giant

protein antigen 332 [Plasmodium falciparum]

Identities = 41/173 (23%), Positives = 87/173 (49%), Gaps = 10/173 (5%)

Query: 1 VSEESKEVEVTKESQTLGLNEAKSMTIGEAVRKQSE---IKAGVIKDDSDILDKYIKQHR 56

+ E + V + KE + GL+ + + ++V +Q+E I + K+ S ++ ++

Sbjct: 78 IEEAENNVWIEKEVEEEGLDNEEVIDEEDSVSEQAEEVYINEEILKESDVEDVKVENE 137

Query: 57 ---DEVSSQKFDKAYTELDASLDNFIIKKQREALSKAGLVDDPEVSAESAQDSTLVVEV 113

+EV+ + + LDN++ ++ E++++ +VD+ P S E E +S ++EE+

Sbjct: 138 LMNEEVNEETQSAENNEEDKELDNYVVEETESVTEEVVVDEVPNSKEVQEIES-IIIEI 196

Query: 114 AEDLAPMETTAVVTGIPVEATVPVLDLDPSEVIPEPQMTKEPKRDQFLSED 166

ED + G +E V + D SE ++ E +T+E K++ ++ED

Sbjct: 197 VEDGLITDDLVGQGGVIEEVVEEVGSD-SEGI VEEASITEVEKKES-VTED 247

An alignment of the GAS and GBS proteins is shown below.

Identities = 234/506 (46%), Positives = 304/506 (59%), Gaps = 36/506 (7%)

Query: 1 MSEDQKHPFFEPKKTGDFKDAKEMTVEEAVRKDSEIKAGITEEDSILDKYIKQHRDE 60

+SE+ K E KE+ L +AK MT+ EAVRK SEIKAG+T++DSILDKYIKQHRDE

Sbjct: 1 VSEESKE--VEVTKESQTLGLNEAKSMTIGEAVRKQSEIKAGVTKDDSDILDKYIKQHRDE 58

Query: 61 VASQKFETKSSDFANLDTASLDDFIKKQRELSAMLAEEELSKLDNSVSEQDTEANAV 120

V+SQKF+ K + LDTASLD+FIKKQRE LS A + + ++ S EQD+

Sbjct: 59 VSSQKFDK---YTELDASLDNFIIKKQREALSK---AGLVDDPEVSAESAQDSTLVVEE 112

Query: 121 SPKESSSQEQENSVTVPVPLNT-----EAEPTATEP--DSTIADSEYKSS 164

++ + E VT +P T E + T EP D +++ + +

Sbjct: 113 VAEDLAPMETTAVVTGIPVEATVPVLDLDPSEVIPEPQMTKEPKRDQFLSEDSHHPAK 172

Query: 165 SKKRGGIVGTLIALILLIVAIPIGYNYFKNNNSTNSQTATSQSSSSKATTSSEEDKKAS 224

+ G + L L+L ++ +FG+N+F +S + S+ + + T S+++ +

Sbjct: 173 QNTKKGWLIALLLAILAVVFGWNHFLRQDSGKTTQTASKQTTSLTQDSAKKATRLK 232

Query: 225 QNLDNFNKSYANFFVDDKKTQLKNSEFDKLSELEKKVDALKGTKYYGKVVKFDSLKRQI 284

F K Y F+ D K++LKNs F L +LE + AL+G+ YY K K K DSLK+ I

Sbjct: 233 AAAKAFKLYGTFYTDATKSLKNSAFATLPDLAALKALEGSAYYDKAKAKVDSLKKAI 292

Query: 285 DAVKAVNDKFKSPAVVDGKKSEKLEVKDGANFDSLDSKTLNTGNASLDSLHHSIVSTGRN 344

A+ AVN KF S VVDG+K EVK ANFD L S TL GNA+LD++L + + GR

-1072-

Sbjct: 293 AAITAVNGKFVSDVVVDGEKVSA-EVKADANFDDLSSATLTIGNANLDAVLQASITEGRQ 351

Query: 345 QVKQSEEQASSNKVSDTQITEQPNVTNGQSSSSAATINNQAAGTAS---GNLERNRSRVP 401
 Q+ E A K ++ Q Q GQS+S A + G S +L+R+ SRVP

5 Sbjct: 352 QLASKAEAA---KAANEQAV-QDQAAQGQSTSVAPS---GYGLTSYDPASLQRHLSRVP 403

Query: 402 YNNAIADTGNPAWIFNPGVLEKIVATSQARGYFSGNNYILEPVNIINGNGYYNMFKLDG 461
 YN IAD NP+W FNPVLEKIVATSQARGY SGN YILEPVNIINGNGYYNMFK DG

10 Sbjct: 404 YNQDVIADRANPSWAFNPGVLEKIVATSQARGYISGNQYILEPVNIINGNGYYNMFKPDG 463

Query: 462 TYLFSINAKTGYFVGNAPGRADSLDY 487
 TYLFSIN KTGYPVGN G AD+LDY

Sbjct: 464 TYLFSINCKTGYFVGNKGYADALDY 489

- 15 SEQ ID 2970 (GBS351) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 73 (lane 2; MW 57kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 81 (lane 5; MW 82kDa).

GBS351-GST was purified as shown in Figure 216, lane 4.

- Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.
- 20

Example 972

A DNA sequence (GBSx1031) was identified in *S.galactiae* <SEQ ID 2973> which encodes the amino acid sequence <SEQ ID 2974>. Analysis of this protein sequence reveals the following:

Possible site: 19

25 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3169(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

30 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2975> which encodes the amino acid sequence <SEQ ID 2976>. Analysis of this protein sequence reveals the following:

Possible site: 19

35 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3169(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

40 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 129/160 (80%), Positives = 149/160 (92%)

45 Query: 1 MTKEVVVESFELDHTIVKAPYVRLISEEVGPVGDITITNFDIRLIQPNENAIDTAGLHTIE 60
 MTKEV+VESFELDHTIVKAPYVRLISEE GP GD ITNFD+RL+QPN+N+I+TAGLHTIE

Sbjct: 1 MTKEVIVESFELDHTIVKAPYVRLISEEFGPKGDRITNFDVRLVQPNQNSIETAGLHTIE 60

50 Query: 61 HLLAKLIRQRINGLIDCSPFGCIRTGFHIMWGKQDATEIAKVIKSSLEA TAGGVTWEDVP 120
 HLLAKLIRQRI+G+IDCSPFGCIRTGFH+IMWGK +T+IAKVIKSSLE IA G+TWEDVP

Sbjct: 61 HLLAKLIRQRIDGMIDCSPFGCIRTGFHIMWGKHSSTDIKVIKSSLEETATGITWEDVP 120

Query: 121 GTTIESCGNYKDHSLSAQEWAKLILSQGISDNAFERHIV 160
 GTT+ESCGNYKDHS L +A+EWA+LI+ QGISD+ F RH++

55 Sbjct: 121 GTTIESCGNYKDHS LFAAKEWAQLIIDQGISDDPF SRHVI 160

-1073-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 973

A DNA sequence (GBSx1032) was identified in *S.galactiae* <SEQ ID 2977> which encodes the amino acid sequence <SEQ ID 2978>. Analysis of this protein sequence reveals the following:

Possible site: 32

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

10 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

15 >GP:AAF34762 GB:AF228345 unknown [Listeria monocytogenes]
 Identities = 302/532 (56%), Positives = 400/532 (74%), Gaps = 14/532 (2%)

Query: 4 IILAMVCALIGLIIGYVAISMKMKSSKEAAELTLNAEQDAVDLRGKAEIEAEHIRKAAE 63
 I + ++ +L+ LI+G V S+ KSS E+ RG AE+ E +K AE
20 Sbjct: 3 IAITIISSLLFLIVGLVVGSLIFKSS-----TEKKLAAARGTAEIIVEDAKKBAE 52

Query: 64 RESKAHQKELLLEAKEEARKYREEIEKEFKSDRQELKQMEARLITDRASSLDRKDNLSNK 123
 +KE LLEAKEE + R EIE E + R E ++ E RL R +LDRKD +LS +
25 Sbjct: 53 TT----KKEALLEAKEENHRLRTEIENELRGRRTETQKAENRLLQREENLDRKDTLSKR 108

Query: 124 EKMLDSKEQSLTDKSRHINEREQEIATLETKKVEELSRIAELSQEAKDIILADTEKDLA 183
 E L+ KE+S++ + + I E+E ++A + + EL RI+ LS+EEAK IIL E++L
30 Sbjct: 109 EATLERKEBSISKRQQIEEKESKLAEMIQAEOQTELERISALSKEBAKSIILNQVEEBLT 168

Query: 184 HDIATRIKEAEREVKDRSNKIAKDLLAQAMQRLAGEYVTEQTITTVHLPDDNMKGRIIGR 243
 HD A +KE+E K+ S+K AK++L+ A+QR A ++V E T++ V LP+D MKGRIIGR
35 Sbjct: 169 HDTAIMVKESENRAKESDKKAKNILSLAIQRCADHVAETTTSVVTLPNDEMKGRIIGR 228

Query: 244 EGRNIRTLESITGIDVIIDDTPEVVVLSGFDPIRREIARMTLES LIQDGRHPARIEELV 303
 EGRNIRTLE+LTGID+IIDDTPE V+LSGFDPIRREIAR+ LE L+QDGRHPARIEE+V
40 Sbjct: 229 EGRNIRTLETLTGIDLIIDDTPEAVILSGFDPIRREIARIALEKLVQDGRHPARIEEMV 288

Query: 304 EKNRLEMDQRIREYGEAAAYEIGAPNLHPDLIKIMGRLQFRTSYQGNVLRHSVEVGKLAG 363
 +K R E+D+ IRE GE A +E+G ++HPDLIKI+GRL++RTSYQGNVL HS+EV KLAG
45 Sbjct: 289 DKARKEVDEHIREVGEQATFEVGIHSIHDPDLIKILGRLRYRTSYQGNVLNHSLEVSKLAG 348

Query: 364 ILAGELGENVDLARRAGFLHDMGKAIDREVEGSHVEIGMEFARKYKEHPVNTIASHHG 423
 ILAGELGE+V LA+RAG LHD+GKAID E+EGSHVEIG+E A KYKE+ +V+N+IASHHG
50 Sbjct: 349 ILAGELGEDVTLAKRAGLLHDIGKAIDHEIEGSHVEIGVELATKYKENDVVINSIASHHG 408

Query: 424 DVEPDSVIAVIVAADALSSARPGARNESMENYIKRLRDLEETANGFEGVQNAFALQAGR 483
 D E SVIAV+VAAADALS+ARPGAR+E++ENYI+RL LEEI+ ++GV+ ++A+QAGR
55 Sbjct: 409 DTEATSVIAVLVAAADALSAARPGARSETLENYIRRLKLEETISESYDGVKESYAIQAGR 468

Query: 484 EIRIMVQPGKVSDDQVIMSHKVKREKIEQNLDYPGNIKVTVIREMRAVDFAK 535
 E+RI+V+P + D ++ +R++IE+ LDYPG+IKVTVIRE RAV++AK
60 Sbjct: 469 EVRIIVEPDTIDDLSSYRLARDIRKRIEELDYPGHIKVTVIRETRAVEYAK 520

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2979> which encodes the amino acid sequence <SEQ ID 2980>. Analysis of this protein sequence reveals the following:

Possible site: 32

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

-1074-

bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

5 The protein has homology with the following sequences in the databases:

>GP:AAF34762 GB:AF228345 unknown [Listeria monocytogenes]
 Identities = 299/534 (55%), Positives = 408/534 (75%), Gaps = 14/534 (2%)

10 Query: 2 VNIILLIVSALIGLILGYALISIRLKSAAEELTLNABQEAVIDRGKAEVDAEHKK 61
 + I + I+S+L+ LI+G + S+ KS+ E++ RG AE+ I +
 Sbjct: 1 MTIATITISSLLFLIVGLVVGSLIFKSS-----TEKKLAAARGTAEI----IVED 46

15 Query: 62 AKRESKANRKELLEAKEEARKYREEIEQEFKSERQELKQLETRLAERSITLDRKDENLS 121
 AK+E++ +KE LLEAKEE + R EIE E + R E ++ E RL +R LDRKD +LS
 Sbjct: 47 AKKEAETTKKEALLEAKEENHRLRTEIENELRGRRTETQKAENRLQREENLDRKDTLS 106

20 Query: 122 SKEKVLDSKEQSLTDKSKHIDERQLQVEKLEEEKKAELEKVAAMTIAEAREVILMETENK 181
 +E L+ KE+S++ + + I+E++ ++ ++ + ++ ELE++A++ EA+ +IL + E +
 Sbjct: 107 KREATLERKEESISKRQQQIEEKESKLAEMIQAQTELERISALSKEEAKSIILNQVEE 166

25 Query: 182 LTHEIATRIRDAERDIKDRVTAKDLLAQAMQRLAGEYVTEQTITSVHLPDDNMKGRII 241
 LTH+ A ++++E K+ + K AK++L+ A+QR A ++V E T++ V LP+D MKGRII
 Sbjct: 167 LTHDTAIMVKESENRAKEESDKKAKNILSLAIQRCADHVAETTVSVVTLPNDEMKGRII 226

30 Query: 242 GREGRNIRTLESITGIDVIIDDTPEVVILSGFDPIRRRIARMTLESILIADGRIHPARIEE 301
 GREGRNIRTLE+LTGID+IIDDTPE VILSGFDPIRRRIAR+ LE L+ DGRIPARIEE
 Sbjct: 227 GREGRNIRTLETITGIDLIIDDTPEAVILSGFDPIRRRIARIALEKLVQDGRIPARIEE 286

35 Query: 302 LIVEKNRLEMDNRIREYGEAAAYEIGAPNLHPDLIKMGRQLQRTSFGQNVLRHSVEVGKL 361
 +V+K R E+D IRE GE A +E+G ++HPDLIKI+GRL++RTS+GQNVL HS+EV KL
 Sbjct: 287 MVDKARKEVDEHIREVGEQATFEVGIHSIHPDLIKILGRLYRTSYGQNVLNHSLEVSKL 346

40 Query: 362 AGILAGELGENVALARRAGFLHDMGKAIDREVEGSHVEIGMEFARKYKEHPVVNTIASH 421
 AGILAGELGE+V IA+RAG LHD+GKAID E+EGSHVEIG+E A KYKE+ VV+N+IASH
 Sbjct: 347 AGILAGELGEDVTAKRAGLLHDIGKAIDHEIEGSHVEIGVELATKYKENDVVINIASH 406

45 Query: 422 HGDVEPDSVIAVLVAAADALSSARPGARNESMENYIKRLRDLEEIATSFQVQNSFALQA 481
 HGD E SVIAVLVAAADALS+ARPGAR+E++ENYI+RL LEEI+ S+DGV+ S+A+QA
 Sbjct: 407 HGDTEATSVIAVLVAAADALSARPGARSETLENYIRLEKLEEISESYDGVESYAIQA 466

50 Query: 482 GREIRIMVQPEKISDDQVILSHKVREKIENNLDPGNIKVTVIREMRAVDYAK 535
 GRE+RI+V+P+ I D L+ +R++IE LDYPG+IKVTVIRE RAV+YAK
 Sbjct: 467 GREVRIIVEPDTIDDLSSYRLARDIRKRIEELDYPGHIKVTVIRETRAVEYAK 520

45 An alignment of the GAS and GBS proteins is shown below.

Identities = 451/535 (84%), Positives = 503/535 (93%)

50 Query: 1 MFNIILAMVICALIGLIIGYVAISMKMKSSKEAAELTLNABQDAVDLRGKAEIEAEHIRK 60
 M NIIL +V ALIGLI+GY IS+++KS+KEAAELTLNABQ+AVD+RGKAE++AEHI+K
 Sbjct: 1 MVNIILLIVSALIGLILGYALISIRLKSAAEELTLNABQEAVIDRGKAEVDAEHKK 60

55 Query: 61 AAERESKAHQKELLEAKEEARKYREEIEQEFKSDRQELKQMEARLTDRASSLDRKDENL 120
 A+RESKA++KELLEAKEEARKYREEIE+EFKS+RQELKQ+E RL +R+ +LDRKDENL
 Sbjct: 61 TAKRESKANRKELLEAKEEARKYREEIEQEFKSERQELKQLETRLAERSITLDRKDENL 120

60 Query: 121 SNKEKMLDSKEQSLTDKSRHINEREQEIATLETKKVEELSRIAELSQEEAKDIILADTEK 180
 S+KEK+LDSKEQSLTDKS+HI+ER+ ++ LE +K EL ++A ++ EA+++IL +TE
 Sbjct: 121 SKEKVLDSKEQSLTDKSKHIDERQLQVEKLEEEKKAELEKVAAMTIAEAREVILMETEN 180

65 Query: 181 DLAHDIA TRIKEAREVVKDRSNKIADLLAQAMQRLAGEYVTEQTITTVHLPDDNMKGRI 240
 L H+IATRI++AER++KDR+ K AKDLLAQAMQRLAGEYVTEQTIT+VHLPDDNMKGRI
 Sbjct: 181 KLTHEIATRIRDAERDIKDRVTAKDLLAQAMQRLAGEYVTEQTITSVHLPDDNMKGRI 240

Query: 241 IGREGNIRTLESITGIDVIIDDTPEVVILSGFDPIRRRIARMTLESILIADGRIHPARIE 300
 IGREGNIRTLESITGIDVIIDDTPEVV+LSGFDPIRRRIARMTLESILI DGRIPARIE
 Sbjct: 241 IGREGNIRTLESITGIDVIIDDTPEVVILSGFDPIRRRIARMTLESILIADGRIHPARIE 300

-1075-

Query: 301 ELVEKNRLEMDQRIREYGEAAAYEIGAPNLHPDLIKMGRLOFRTSYGQNVLRHSVEVGK 360
 ELVEKNRLEMD RIREYGEAAAYEIGAPNLHPDLIKMGRLOFRTS+GQNVLRHSVEVGK
 Sbjct: 301 ELVEKNRLEMDNRIREYGEAAAYEIGAPNLHPDLIKMGRLOFRTSFGQNVLRHSVEVGK 360
 5
 Query: 361 LAGILAGELGENVDLARRAGFLHDMGKAIDREVEGSHVEIGMEFARKYKEHPIVVNTIAS 420
 LAGILAGELGENV LARRAGFLHDMGKAIDREVEGSHVEIGMEFARKYKEHP+VVNTIAS
 Sbjct: 361 LAGILAGELGENVALARRAGFLHDMGKAIDREVEGSHVEIGMEFARKYKEHPVVVNTIAS 420
 10
 Query: 421 HHGDVEPDVSVIAVIVAAADALSSARPGARNESMENYIKRLRDLEEIANGFEGVQNAFALQ 480
 HHGDVEPDVSVIAV+VAAADALSSARPGARNESMENYIKRLRDLEEIA F+GVQN+FALQ
 Sbjct: 421 HHGDVEPDVSVIAVLVAAADALSSARPGARNESMENYIKRLRDLEEIATSFQVQNSFALQ 480
 15
 Query: 481 AGREIRIMVQPGKVSDDQVVMISHKVREKIEQNLDYPGNIKVTIVIREMRAVDFAK 535
 AGREIRIMVQP K+SDDQVVI+SHKVREKIE NLDYPGNIKVTIVIREMRAVD+AK
 Sbjct: 481 AGREIRIMVQPEKISDDQVVLISHKVREKIENLDYPGNIKVTIVIREMRAVDYAK 535

SEQ ID 2978 (GBS86) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 7 (lane 6; MW 59kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 13 (lane 5; MW 84kDa).
 20

GBS86-GST was purified as shown in Figure 192, lane 3.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 974

25 A DNA sequence (GBSx1033) was identified in *S.agalactiae* <SEQ ID 2981> which encodes the amino acid sequence <SEQ ID 2982>. Analysis of this protein sequence reveals the following:

Possible site: 28
 >>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4984(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

35 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 975

40 A DNA sequence (GBSx1034) was identified in *S.agalactiae* <SEQ ID 2983> which encodes the amino acid sequence <SEQ ID 2984>. Analysis of this protein sequence reveals the following:

Possible site: 37
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -2.87 Transmembrane 146 - 162 (146 - 162)

45 ----- Final Results -----
 bacterial membrane --- Certainty=0.2147(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

50

-1076-

A related GBS nucleic acid sequence <SEQ ID 8697> which encodes amino acid sequence <SEQ ID 8698> was also identified. Analysis of this protein sequence reveals the following:

```

5  Lipop: Possible site: -1  Crend: 9
   McG: Discrim Score: -10.72
   GvH: Signal Score (-7.5): -5.66
       Possible site: 29
   >>> Seems to have no N-terminal signal sequence
   ALOM program count: 1 value: -2.87 threshold: 0.0
10  INTEGRAL Likelihood = -2.87 Transmembrane 138 - 154 ( 138 - 154)
   PERIPHERAL Likelihood = 3.76 51
   modified ALOM score: 1.07

   *** Reasoning Step: 3

15  ----- Final Results -----
       bacterial membrane --- Certainty=0.2147(Affirmative) < succ>
       bacterial outside --- Certainty=0.0000(Not Clear) < succ>
       bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

20  The protein has homology with the following sequences in the GENPEPT database.

   >GP:AAG21390 GB:AF302051 ABC transporter ATP binding subunit
       [Bacillus licheniformis]
       Identities = 84/218 (38%), Positives = 138/218 (62%), Gaps = 1/218 (0%)

25  Query: 12 DIIKVDHIFKSIGQKTILEDISFSIASNQCVLIGPAGKTTLMSTLLGDISISSGSIT 71
       +++ ++ K+ QKT ++ I FSI + VA++GPNAGAKTT +S +LG + ++G++T
   Sbjct: 3 NVVSLINVTKTFRQKTAVDQIDFSIKKGEIVAILGPNAGAKTTTISMLIGLLKPTAGNIT 62

   Query: 72 IFNLPAHHNRLKYKVAILPQE-NVLPSKFTVRELIDFQRCLEFPEVLPMSLILDYQWSDT 130
       +F+ H R++ K+ + QE +V+P E+I+ R +P+ L + +D
30  Sbjct: 63 LFDSMPHEKRVREKIGTMLQEVSVMPGLRCRVEIIELIRSYYPKPLSFQKLRITLTGLTDK 122

   Query: 131 HLQQFTETLSGGQKRLLAFLVLTGVGKPLLFLDEPTSGMDTSTRQRFWELIATLKKEGVT 190
       L+ E LSGGQKR L F L L G P+L+ DEPT GMD ++R RFW+ + +L ++G T
35  Sbjct: 123 DLKTQAEKLSGGQKRRLLGFALALAGDPELMIFDEPTVGMDITSRNRFWQTVQSLAEQGKT 182

   Query: 191 IVYSSHYIEVEHTADRILVLHKGKLLRDTTPLCHEAR 228
       I++S+HY++E + A RIL+ GK++ D TPL ++R
40  Sbjct: 183 IIFSTHYLQEADDAQRILLFKDGKIVADGTPLQIKSR 220

```

There is also homology to SEQ ID 686.

SEQ ID 8698 (GBS350) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 72 (lane 13; MW 28.9kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 81 (lane 4; MW 54kDa).

GBS350-GST was purified as shown in Figure 226, lane 7.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 976

A DNA sequence (GBSx1035) was identified in *S.agalactiae* <SEQ ID 2985> which encodes the amino acid sequence <SEQ ID 2986>. Analysis of this protein sequence reveals the following:

```

50  Possible site: 35
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
55  bacterial cytoplasm --- Certainty=0.2913(Affirmative) < succ>

```

-1077-

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

- 5 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 977

- 10 A DNA sequence (GBSx1036) was identified in *S.agalactiae* <SEQ ID 2987> which encodes the amino acid sequence <SEQ ID 2988>. Analysis of this protein sequence reveals the following:

Possible site: 31

>>> Seems to have an uncleavable N-term signal seq

- 15 INTEGRAL Likelihood = -10.51 Transmembrane 222 - 238 (214 - 241)
 INTEGRAL Likelihood = -6.90 Transmembrane 104 - 120 (101 - 125)
 INTEGRAL Likelihood = -5.84 Transmembrane 140 - 156 (138 - 159)
 INTEGRAL Likelihood = -5.20 Transmembrane 19 - 35 (18 - 41)
 INTEGRAL Likelihood = -1.28 Transmembrane 164 - 180 (164 - 180)

----- Final Results -----

- 20 bacterial membrane --- Certainty=0.5203(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

- 25 >GP:CAB69806 GB:AJ243712 YVFS protein [Bacillus cereus]
 Identities = 73/239 (30%), Positives = 127/239 (52%), Gaps = 4/239 (1%)
- Query: 9 KMEFLITKRQLANLIMAIGMPVAFFLFFSGFMGEGLTKAIEAIYVRNYMITMAGFSSLSF 68
 K+E L T R + ++ MPV F+ F+ + + +Y+I+MA FS +
 30 Sbjct: 4 KIEILRTFRNKLFIFFSLLMPVMFYIIFTNVVQ--VPQNGDAWKAHYLISMATFSIVGT 60
- Query: 69 AFFTFFFSMKDDQLSNRMQLLRHSFVPMWQYYLAKIIRILFYICLAITVVFLTGHILRQV 128
 A F+F + ++ LL+ +P+P Y AKII +I V+F+ G ++ V
 35 Sbjct: 61 ALFSFGVRLSQERGQGWTHLLKITPLPEGAYLTAKTIAQTVVNAFSILVIFTAGILINHV 120
- Query: 129 SMPIEQWMQSFLLLLGGGATCFIPGLLVSFYFKNTELSMVMANICYMSLAVLGGMWMPITM 188
 + I QW+ + L LL G T F+ G ++ K + + +ANI MSIA++GG+WMPI +
 Sbjct: 121 ELTIQWIGAGLWLLLGVTPLALGTVIGSIGKADAAAGLANILNMSLAIVGGLWMPIEV 180
- 40 Query: 189 FPKWLQALSKLTPTYHLTQVILSPFANSFAGF-SLIILIGYGIIMLVIAVLLSQKRHSI 246
 FPK L+ + + TPTYH A G+ ++ +L GY +I +V++ + +++ ++
 Sbjct: 181 FPKILRTIGEWTPTYHFGSGAWDIVAGKSIGWENIAVLGGYFLIFVVSIIYIRKRQEAV 239

There is also homology to SEQ ID 682 and to SEQ ID 1628.

- 45 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 978

- 50 A DNA sequence (GBSx1037) was identified in *S.agalactiae* <SEQ ID 2989> which encodes the amino acid sequence <SEQ ID 2990>. This protein is predicted to be histidine kinase. Analysis of this protein sequence reveals the following:

Possible site: 49

>>> Seems to have a cleavable N-term signal seq.

INTEGRAL Likelihood = -7.43 Transmembrane 105 - 121 (102 - 124)

-1078-

INTEGRAL Likelihood = -6.95 Transmembrane 130 - 146 (129 - 149)

----- Final Results -----

5 bacterial membrane --- Certainty=0.3972(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9537> which encodes amino acid sequence <SEQ ID 9538> was also identified.

10 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB54584 GB:AJ006400 histidine kinase [Streptococcus pneumoniae]
Identities = 138/350 (39%), Positives = 212/350 (60%), Gaps = 3/350 (0%)

15 Query: 11 MYFIPLVFLIYPIGGILYYHYPFWTLFFTLAFVGAYLYSVIIRGESKYHMIWSTMLTYI 70
M++I L+F+I+PI ++ W L + FV AYL V+ + + W MLTY+
Sbjct: 11 MFWISLIFMIFPILSVVTGWLAWHLLIDILFVVAYL-GVLTTSQRLSWLYWGLMLTYV 69

20 Query: 71 FYMTIFINSGFIWYIYFLSNLLVYRFRDK-LKSEFRFISFACTLATVVF-LCFFKASDFGD 128
T F+ +IW+ +FLSNLL Y F + LKS +F VV L F+ +
Sbjct: 70 VGNTAFVAVNYIWFFFFLSNLLSYHFSVRSLSLHVWTFLLAQVLVVGQLLIFQRIEVEF 129

25 Query: 129 RIMFLIVPIFCIGYMWIAIENRNSEEQREKIAEQNQYINILSAENERNRIGRDLHDSLGH 188
L++ F + + R E+ +E +QN IN+L AENER+RIG+DLHDSLGH
Sbjct: 130 LFYLLVILTFVDLMTFGLVIRIVEDLKEAQVKQNAQINLLLAENERSRIGQDLHDSLGH 189

30 Query: 189 TFAMTLKTELALKLEKRNYPKVQKELSELNHISHQSMSEVRQIVSNLKYRTVVEEIDE 248
TFAM+++KT+LAL+L + Y +V+KEL E++ IS SM+EVR IV NLK RT+ E++
Sbjct: 190 TFAMLSVKTDLALQLFQMEAYPQVEKELKEIHQISKDSMNEVRTIVENLKSRTLTSLT 249

35 Query: 249 LYRLFQLSNIKLTVVNKLETSQSPVTQSTITMILKELSNNIVKHAEADSVELSLVRQGA 308
+ ++ +++ I++ V N L+ S L+ +ST +MIL EL NI+KHA+A V L L R
Sbjct: 250 VKKMLELAGIEVQVENHLDKSSLTQLESTASMILLELVTNLIKAKASKVYVYKLERTK 309

Query: 309 TINIEMIDNGCGFTNLDGDELHSIQERLTIVEGTLTILSRSKPHTHIQVVL 358
+ + + D+GCGF ++ GDELH+++ R+ G ++++S+ PT +QV L
Sbjct: 310 ELILTVRDDGCGFASISGDELHTVRNRVFPFSGEVSVISQKHPTEVQVRL 359

There is also homology to SEQ ID 2992.

40 A related GBS gene <SEQ ID 8699> and protein <SEQ ID 8700> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 8
McG: Discrim Score: 10.90
GvH: Signal Score (-7.5): -2.42
Possible site: 49

45 >>> Seems to have a cleavable N-term signal seq.
ALOM program count: 2 value: -7.43 threshold: 0.0
INTEGRAL Likelihood = -7.43 Transmembrane 105 - 121 (102 - 124)
INTEGRAL Likelihood = -6.95 Transmembrane 130 - 146 (129 - 149)
50 PERIPHERAL Likelihood = 0.16 61
modified ALOM score: 1.99

*** Reasoning Step: 3

----- Final Results -----

55 bacterial membrane --- Certainty=0.3972(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1079-

Example 979

A DNA sequence (GBSx1038) was identified in *S.agalactiae* <SEQ ID 2993> which encodes the amino acid sequence <SEQ ID 2994>. This protein is predicted to be response regulator. Analysis of this protein sequence reveals the following:

```

5   Possible site: 28
    >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -0.16    Transmembrane    49 - 65 ( 49 - 65)

10  ----- Final Results -----
      bacterial membrane --- Certainty=0.1065(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15  >GP:CAB54585 GB:AJ006400 response regulator [Streptococcus pneumoniae]
    Identities = 95/153 (62%), Positives = 125/153 (81%), Gaps = 3/153 (1%)

    Query: 1   MKLLVAEDQSMRLDAMCQLLLMEERSVSTIDQAGNGGEAIAILSNKAIDVAILDVEMPILS 60
              MK+LVAEDQSMRLDAMCQLL+++ V ++ QA NG EAI +L +++D+AILDVEMP+ +
20  Sbjct: 1   MKVLVAEDQSMRLDAMCQLMLQPDVESVFQAQNGQEAIIQLLEKESVDIAILDVEMPVKT 60

    Query: 61   GLDVLEWVRKYQ-NVKVVIIVTTFKRSGYFQRAIRSNVDAYVLKDRSVADLMKTIQKVLGS 119
              GL+VLEW+R + KV++VTTFKR GYF+RA+++ VDAYVLK+R++ADLM+T+ VL G
25  Sbjct: 61   GLEVLEWIRAEKLETKVVVVTTFKRPGYFERAVKAGVDAYVLKERNIADLMQTLHTVLEG 120

    Query: 120  GKEYSPELMENVI--SNPLSEQEIKILSLIAQG 150
              KEYSPELME V+ NPL+EQEI +L IAQG
30  Sbjct: 121  RKEYSPELMEVMMHPNPLTEQEIAVLKGIAQG 153

```

30 There is also homology to SEQ ID 2996.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 980

35 A DNA sequence (GBSx1039) was identified in *S.agalactiae* <SEQ ID 2997> which encodes the amino acid sequence <SEQ ID 2998>. Analysis of this protein sequence reveals the following:

```

    Possible site: 34
    >>> Seems to have an uncleavable N-term signal seq

    INTEGRAL    Likelihood = -6.69    Transmembrane    158 - 174 ( 145 - 184)
    INTEGRAL    Likelihood = -4.94    Transmembrane    11 - 27 ( 8 - 31)
40  INTEGRAL    Likelihood = -3.93    Transmembrane    74 - 90 ( 73 - 92)
    INTEGRAL    Likelihood = -2.39    Transmembrane    103 - 119 ( 102 - 119)
    INTEGRAL    Likelihood = -1.86    Transmembrane    42 - 58 ( 38 - 59)

45  ----- Final Results -----
      bacterial membrane --- Certainty=0.3675(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

50  >GP:AAB85965 GB:AE000909 unknown [Methanothermobacter
    thermoautotrophicus]
    Identities = 46/183 (25%), Positives = 81/183 (44%), Gaps = 11/183 (6%)

    Query: 5   KERFDTLSDAILAIAMTILVLEI-----KTPATMGDIGDFTRNIGLFIVSFVVVFNF 57
              K+R + L DAI AIAMTILVL I PA I ++ + +SF+++ FW
55  Sbjct: 6   KKRLEGLVDAIFAIAMTILVLGIDVPTGTMSVPAMDAYIMGLASDLYSYCLSFLLLGVF 65

```

-1080-

5 Query: 58 YERAQNSLDAQKTNDDEIIALDIEHLGICLIPLFTKFMISFENHNFAVMAYGLLTLLVGL 117
 + + +K + I ++I+ + + L+P TK ++ + + + L L +GL
 Sbjct: 66 WVNHHMFEEKLEKVDGTGFIWINIVWLMVVVLVPFSTKLTGNYGDLVTPNLFHLNMLTIGL 125

10 Query: 118 TSDIIRIRLASDYDLVTIPSELKERVIVMTTFAIRSVVVRFTIIILAYFLPEVGIFAYLV 177
 + I L+ I ++K + ++ + +IL PE AY V
 Sbjct: 126 LLSMSWIYTQRNGLMDIGENEYRLILKKNLLMPLAAI----LALILTPIAPEYSSTAYAV 181

10 Query: 178 IPL 180
 + L
 Sbjct: 182 LIL 184

No corresponding DNA sequence was identified in *S.pyogenes*.

- 15 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 981

- 20 A DNA sequence (GBSx1040) was identified in *S.agalactiae* <SEQ ID 2999> which encodes the amino acid sequence <SEQ ID 3000>. This protein is predicted to be guanylate kinase (gmk). Analysis of this protein sequence reveals the following:

Possible site: 16
 >>> Seems to have an uncleavable N-term signal seq

25 ----- Final Results -----
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

30 >GP:CAB13441 GB:Z99112 similar to guanylate kinase [Bacillus subtilis]
 Identities = 121/202 (59%), Positives = 155/202 (75%)

35 Query: 1 MSERGLLIVFSGPSGVGKGTVRQEIFSTPDHKFDYSVSMTTTRPQRPGEVDGVDYFFRTRE 60
 M ERGLLIV SGPSGVGKGTVRQ IFS D KF+YS+S+TTR R GEV+GVDYFF+TR+
 Sbjct: 41 MKERGLLIVLSGPSGVGKGTVRQAIIFSQEDTKFEYSISVTTTSPRGEVNGVDYFFKTRD 100

40 Query: 61 EFEALIKEGQMLEYAEYVGNYYGTPLSYVNETLDKGIDVFLEIEVQGALQVKSQVPDGVF 120
 EFE +I + ++LE+AEYVGNYYGTP+ YV +TL G DVFLEIEVQGALQV++ P+G+F
 Sbjct: 101 EFEQMIADNKLLEWAEYVGNYYGTPVDYVQTLQDGKDVFLIEVQGALQVRNAFPEGLF 160

45 Query: 121 IFLTPPDLEELEERLVGRGTDSPVIAQRIERAKEEIALMREYDYAVVNDQVSLAAERVK 180
 IFL PP L EL+ R+V RGT++ +I R++ AK EI +M YDY V ND V A +++K
 Sbjct: 161 IFLAPPSELKKNRIVTRGTETDALIENRMKAAKAEIEMMDAYDYVVENDNVETACDKIK 220

45 Query: 181 RVIEAEHYRVDRVIGRYTNMVK 202
 ++ AEH + +RV RY M++
 Sbjct: 221 AIVLAEHLKRERVAPRYKKMLE 242

- 50 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3001> which encodes the amino acid sequence <SEQ ID 3002>. Analysis of this protein sequence reveals the following:

Possible site: 16
 >>> Seems to have an uncleavable N-term signal seq

55 ----- Final Results -----
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

-1081-

The protein has homology with the following sequences in the databases:

>GP:CAB13441 GB:Z99112 similar to guanylate kinase [Bacillus subtilis]
Identities = 123/203 (60%), Positives = 157/203 (76%)

5 Query: 1 MSERGLLIVFSGPSGVGKGTVRQEIFSTPDHKEFYSVSMTTRPQRPGEVDGVDYFFRTRE 60
M ERGLLIV SGPSGVGKGTVRQ IFS D KFEYS+S+TTR R GEV+GVDYFF+TR+
Sbjct: 41 MKERGLLIVLSGPSGVGKGTVRQAIIFSQEDTKFEYSISVTTTRSPREGEVNGVDYFFKTRD 100

10 Query: 61 EFEELIKTGQMLEYAEYVGNYYGTPLTYVNETLDKGIDVFLEIEVQGALQVKSVPDGVF 120
EFE++I ++LE+AEYVGNYYGTP+ YV +TL G DVFLEIEVQGALQV++ P+G+F
Sbjct: 101 EFEQMIADNKLLEWAEYVGNYYGTPVDYVEQTLQDGKIDVFLEIEVQGALQVRNAFPEGFLF 160

15 Query: 121 VFLTPPDLEDELDRLVGRGTDSEVIAQRIERAKEEIALMREYDYAVVNDEVALAAERVK 180
+FL PP L EL++R+V RGT++ +I R++ AK EI +M YDY V ND V A +++K
Sbjct: 161 IFLAPPSELSELKNRIVTRGTETDALIENRMKAAKAEIEMMDAYDYVENDNVETACDKIK 220

20 Query: 181 RIIETEHRFRVERVIGRYDKMIKI 203
I+ EH + ERV RY KM+++
Sbjct: 221 AIVLAEHLKRERVAPRYKKMLEV 243

An alignment of the GAS and GBS proteins is shown below.

Identities = 186/204 (91%), Positives = 197/204 (96%)

25 Query: 1 MSERGLLIVFSGPSGVGKGTVRQEIFSTPDHKEFYSVSMTTRPQRPGEVDGVDYFFRTRE 60
MSERGLLIVFSGPSGVGKGTVRQEIFSTPDHKEFYSVSMTTRPQRPGEVDGVDYFFRTRE
Sbjct: 1 MSERGLLIVFSGPSGVGKGTVRQEIFSTPDHKEFYSVSMTTRPQRPGEVDGVDYFFRTRE 60

30 Query: 61 EFEALIKGQMLEYAEYVGNYYGTPLSYVNETLDKGIDVFLEIEVQGALQVKSVPDGVF 120
EFE LIK GQMLEYAEYVGNYYGTPL+YVNETLDKGIDVFLEIEVQGALQVKSVPDGVF
Sbjct: 61 EFEELIKTGQMLEYAEYVGNYYGTPLTYVNETLDKGIDVFLEIEVQGALQVKSVPDGVF 120

35 Query: 121 IFLTPPDLEELERLIVGRGTDSEVIAQRIERAKEEIALMREYDYAVVNDQVSLAAERVK 180
+FLTPPDL+ELE+RLVGRGTDSEVIAQRIERAKEEIALMREYDYAVVND+V+LAAERVK
Sbjct: 121 VFLTPPDLEDELDRLVGRGTDSEVIAQRIERAKEEIALMREYDYAVVNDEVALAAERVK 180

Query: 181 RVIEAEHYRVDRVIGRYTNMVKET 204
R+IE EH+RV+RVIGRY M+K T
Sbjct: 181 RIIETEHRFRVERVIGRYDKMIKIT 204

40 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 982

A DNA sequence (GBSx1041) was identified in *S.agalactiae* <SEQ ID 3003> which encodes the amino acid sequence <SEQ ID 3004>. Analysis of this protein sequence reveals the following:

45 Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
50 bacterial cytoplasm --- Certainty=0.1763(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

55 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3005> which encodes the amino acid sequence <SEQ ID 3006>. Analysis of this protein sequence reveals the following:

Possible site: 35
>>> Seems to have no N-terminal signal sequence

-1082-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1551(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

5

An alignment of the GAS and GBS proteins is shown below.

Identities = 95/105 (90%), Positives = 100/105 (94%), Gaps = 1/105 (0%)

10

Query: 1 MMLKPSIDTLLDKVPSKYSVLVILQAKRAHELEAGEKATQDFKSVKSTLRALEEIESGNV 60
 MMLKPSIDTLLDKVPSKYSVLVILQAKRAHELEAG TQ+FKSVKSTL+ALEEIESGNV
 Sbjct: 1 MMLKPSIDTLLDKVPSKYSVLVILQAKRAHELEAGATPTQEFKSVKSTLQALEEIESGNV 60

15

Query: 61 IHPDPSAKRASVRARIEAERLAKEEEERKIKEQIAKEK-EDGEKI 104
 IHPDPSAKR +VRA+IEAERLAKEEEERKIKEQIAKEK E+GEKI
 Sbjct: 61 IHPDPSAKREAVRAKIEAERLAKEEEERKIKEQIAKEKEEBEGEKI 105

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 983

20 A DNA sequence (GBSx1043) was identified in *S.agalactiae* <SEQ ID 3007> which encodes the amino acid sequence <SEQ ID 3008>. Analysis of this protein sequence reveals the following:

Possible site: 24
 >>> Seems to have no N-terminal signal sequence

25

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3413(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

30

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB13444 GB:Z99112 primosomal replication factor Y (primosomal protein N') [Bacillus subtilis]
 Identities = 377/807 (46%), Positives = 529/807 (64%), Gaps = 21/807 (2%)

35

Query: 6 AQVIVDIPLMQTDKPFYSYAIPKDLQVQGVVRVHVPFGRGNRLQLQGFVVGFRDDDELET 65
 A+VIVD+ D+PF Y IP L+ +++ G+RV VPFG R +QGFV ++ +L
 Sbjct: 4 AEVIVDVSTKNIDRPFYKIPDHLKGMIKTGMRVIVVPFGP--RKIQGFVTAVKEASDLG 61

40

Query: 66 KDIAEV---LDPEPVLNQEQDLADQMRHTVFSYKISILKSMPLSLNSQYDKLLL---A 119
 K + EV LD PVL +E + L+ + S+KI+ I++MLP+ I ++Y+K L
 Sbjct: 62 KSVKEVEDLLDLPVLTTELMILSSWLSDKTSLFKITALQAMLPAALKAKYEKELKIAHG 121

45

Query: 120 TDTLPSEDREDLFGHKTEIVFSSLSQDAKKA--GRLIQKGFIEVQYLAQDKKTIKTEKIY 178
 D P +R LF +++S + + K R +QKG I+V Y K K +
 Sbjct: 122 ADLPPQVER--LFSETKTLLYSDIPDHETLKLIRHVQKGDIDVTYKVAQKTNNKMRVRI 179

50

Query: 179 KINRTLLEKSQ---IAARAKRLELKEFLENPOPGRLTALN---KQFSSPVVNFRE 230
 + N + E ++ ++ +A K+ + FL+ P+ ++ A SS + +
 Sbjct: 180 QANASKEELAKQAEGLSRQAQKQAILHFLISEPEGVKIPAAELCKKTDTSATIKTLIQ 239

55

Query: 231 EGIIEVIEKEASRSNYFKGILKTDFLDLNQEQAKVVVQIGKEQNKPFLLLEGITGS 290
 +G+++ +E R K KT+ L L EQ + + + + ++K FLL G+TGS
 Sbjct: 240 KGLLKESYEEVYRDPYQDKMFKKTEPLPLTDEQRAAFPIRETLDSEHKVFLHGVITGS 299

60

Query: 291 GKTEVYLHIIDNVKLKGTAVLVPEISLTPQMTNRFISRFKQVAIMHSGSLSEGEKFDE 350
 GKTE+YL I+ VL GK AIVLVPEISLTPQM NRF RFG QVA+MHSGLS GEK+DE
 Sbjct: 300 GKTEIYLSIEKVLAKGKEAIVLVPEISLTPQMVNRFKGRFGSQVAVMHSGSLTGEKYDE 359
 Query: 351 WRKIKSGQAKVVVGARSAIFAPLENIGAIIDEEHESTYKQESNPRYHARDVALLRAEYY 410
 WRKI + ++VVGARSAIFAP EN+G IIDEEHES+YKQE PRYHA++VA+ RAE++
 Sbjct: 360 WRKIHRKEVRLVVGARSAIFAPPENLGMIIIDEEHESYKQEMPRYHAKEVAIKRAEHH 419

-1083-

- Query: 411 KAVLLMGSA TPSTIESRARASRDVYKFLELKHANPKARIPQVEIIDFRNFIGQQEVSNFT 470
 +++GSATP++ES ARA + VY+ L LKHR N + +P+V ++D R + S F+
 Sbjct: 420 SCPVVLGSATPTLESYARAQKGVYELLSLKHVRNHRV-MPEVSLVDMREELRNGNRSMF 478
- Query: 471 SYLLDKIRDRDLKKEQVVLMLNRRGYSSFIMCRDCGYVDQCPNCDISLTLHMATKTMNCH 530
 L++K+ + + K EQ VL LN+RGYSSF+MCRDCGYV QCP+CDIS+T H + + CH
 Sbjct: 479 VELMEKLEETIAKGEQAVLFLNKRGYSSFVMCRDCGYVPQCPHCDISMTYHRYGQRLKCH 538
- Query: 531 YCGFEKPIPRTPCNCNSKSISSYYGTGTQKAYEELLKVIIPDAKILRMDVDTTRQKGGHESI 590
 YCG E+P+P TCP C S+ I ++GTGTQ+ EEL KV+P A+++RMDVDTT +KG HE +
 Sbjct: 539 YCGHEEPVPHTCPECASEHIRFFGTGTQRVEEELTKVLPSARVIRMDVDTTSRKGAEKL 598
- Query: 591 LKRFGNHEADILLGTQMIAGLDFFENVTLVGVLNADTSLNLPDFRSSERTFQLLTQVAGR 650
 L FG +ADILLGTQMIAGLDFFENVTLVGVL+ADT+L++PDRS+E+TFQLLTQV+GR
 Sbjct: 599 LSAFGEKGADILLGTQMIAGLDFFENVTLVGVLADTTLHIPDFRSAEKTFFQLLTQVSGR 658
- Query: 651 AGRAEKEGEVVIQTYNPNHYAIIQLAQKQDFEAFYQYEMNIRRLGYPPYYFTVGLTLSHK 710
 AGR EK G V+IQTY P+HY+IQL + D+E FYQ+EM RR+ YPPYY+ +T+SH+
 Sbjct: 659 AGRHEKPGHVIIQTYTPSHYSIQLTKTHDYETFYQHEMAHRREQSYPPYYLALVTSHE 718
- Query: 711 DEEWLIRKSYEVLSELLKQGFSDKVKLLGPTPKPIARTHNLVHYQIIKRYFEDNLELVIN 770
 + + ++ LK K+LGP+ PIAR + Y YQ +IKY+ E L +L
 Sbjct: 719 EVAKAAVTAEKIAHFLKANGADTKILGPSASPIARIKDRYRYQCVIKYQETQLSALLK 778
- Query: 771 RLLD-MTQDKENRDLRLAIDHEPQNM 796
 ++L+ ++ E + + ++ID P MM
 Sbjct: 779 KILEHYKREIEQKHVMISIDMNPYMM 805
- 30 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3009> which encodes the amino acid sequence <SEQ ID 3010>. Analysis of this protein sequence reveals the following:
- Possible site: 32
 >>> Seems to have no N-terminal signal sequence
- 35 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1396(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
- 40 An alignment of the GAS and GBS proteins is shown below.
- Identities = 556/793 (70%), Positives = 659/793 (82%), Gaps = 1/793 (0%)
- Query: 4 KLAQVIVDIPLMQTDKPFSSYAIPKDLQVQGVVRVHPFGRGNRLQLQGFVVGFRDDDEL 63
 K+A VIVDIPLMQTDKPFSSY IPK+L LVQ+G RVHVPFG+GNRLQLQGF++GF +D
 Sbjct: 12 KVAHVIVDIPLMQTDKPFSSYGIPKELVSLVQLGSRVHVPFGKGNRLQLQGFIIIGFGQEDSS 71
- Query: 64 ETKDIAEVLDFEPVLNQEQQLDLADQMRHTVFSYKISILKSMPLSLNSQYDKLLLATDTL 123
 K I VLD EPVLNQEQQL IADQ+R TVFSYKI++LK+M+P+LLNS YDK+L L
 Sbjct: 72 SLKLIQTVDLPEPVLNQEQQLTLADQLRKTVFSYKITLLKAMIPNLLNSNYDKVLRPESGL 131
- Query: 124 PSEDREDLFHGKTEIVFSSLSQDAKKAGRLIQKFIEVQYLAKDKKTIKTEKIYKINRT 183
 DR+ LF K +++S+L + K A + IQ G I V YLAKDKK +KTEK Y ++
 Sbjct: 132 KKSDDRDLFEGKPSVLYSTLDREKEKIALKGIIQAGHITVSYLAKDKKNLKTEKYHVDLD 191
- Query: 184 LLEKSQIAARAKRLELKEFLLENPQPGRLTALNKQFSSPVVNFREEGIEVIEKEASR 243
 L I++RAKKR LK++LL + + +L L + FS VV +F +I + E+ R
 Sbjct: 192 ALAVHPISSRAKKRQLKDYLLTHTKEAKLATLYQAFSRDVVAYFVINHLIRIDERPIDR 251
- Query: 244 SDNYFKGILKTDFLDLNQEQAKVVKIVVDQIGKEQNKPFLLLEGITGSGKTEVYLHIIDNV 303
 S++YF I + FL LN++QA V +V+QIGK +KPFL+EGITGSGKTEVYLHII+ V
 Sbjct: 252 SESYFDQIKPSSFLTLNEQQA SAVTEIVEQIGKP-SKPFLIEGITGSGKTEVYLHIEAV 310
- Query: 304 LKLGTAVILVPEISLTPQMTNRFISRFKQVAIMHSGLSGEKGFDEWRKIKSGQAKVVV 363
 LK KTAIVLVPEISLTPQMT+RFISRFKQVAIMHSGLS+GEKGFDEWRKIK+GQAKVVV
 Sbjct: 311 LKQDKTAVILVPEISLTPQMTSRFISRFKQVAIMHSGLSGEGKGFDEWRKIKTGQAKVVV 370

-1084-

Query: 364 GARSAlFAPLENIGAlIIDEHESTYKQESNPRYHARDVALLRAEYYKAVLLMGSATPSI 423
 GARSAlF+PLE IGAlIIDEHESTYKQESNPRYHAR+VALLRA+++AV++MGSATPSI
 Sbjct: 371 GARSAlFSPLERIGAlIIDEHESTYKQESNPRYHAREVALLRAKHQA VVVV MGSATPSI 430

Query: 424 ESRARASRDVYKFLELKHANPKARIPQVEIIDFRNFIGQQEVSNTSYLLDKIRDRLDK 483
 ESRARAS+ VY F++L RANP A+IP+V I+DFR++IGQQ VSNFT YL+DKI++RL K
 Sbjct: 431 ESRARASKGVYHFIQLTQRANPLAKIPEVTIVDFRDYIGQQAVSNFTPYLIDIKIKERLVK 490

Query: 484 KEQVVLMLNRRGYSSFIMCRDCGYVDQCPNCISLTLHMAKTMTNCHYCGFEKPIPTCP 543
 KEQVVLMLNRRGYSSF+MCRDCGYVD+CPNCISLTLHM TKTMTNCHYCGF+KPIP TCP
 Sbjct: 491 KEQVVLMLNRRGYSSFVMCRDCGYVDKCPNCISLTLHMDTKMTNCHYCGFQKPIPTCP 550

Query: 544 NCNSKSISYYGTGTQKAYEELLKVIPDAKILRMDVDITTRQKGGHESILKRFNGHEADILL 603
 C+S SI YYGTGTQKA++EL VIP+AKILRMDVDITR+K H++IL FG EADILL
 Sbjct: 551 ECHSNSIRYYGTGTQKAFDELQGVPEAKILRMDVDITRKKRSHKTILDSFGRQEADILL 610

Query: 604 GTQMIAGLDFFPNVTLVGVLNADTSLNLPDFRSSERTFQLLTQVAGRAGRAEKEGEVVIQ 663
 GTQMIAGLDFFPNVTLVGVLNADTSLNLPDFR+SE+TFQLLTQVAGRAGRA K GEV+IQ
 Sbjct: 611 GTQMIAGLDFFPNVTLVGVLNADTSLNLPDFRASEKTQQLLTQVAGRAGRAHKPGEVLIQ 670

Query: 664 TYNPNHYAIQLAQKQDFEAFYQYEMNIRRLQGYPPYYFTVGLTSLHKDEEWLIRKSYEVL 723
 TYNP+HYAIQLA+KQDFEAFY+YEM+IR Q+ YPPYYFTVG+TLSH+ E +++K+Y+V
 Sbjct: 671 TYNPDHYAIQLAKKQDFEAFYRYEMSIRHQMAYPPYYFTVGITLSHRLEASVVKAYQVT 730

Query: 724 SLLKQGFSDKVKLLGPTPKPIARTHNLYHYQIIKYRFEDNLELVNRLLDMTQDKENRD 783
 LLK SD +K+LGPTPKPIARTHNLYHYQI++KYRFEDNLE LNR+LD +Q+ +NR
 Sbjct: 731 ELLKSHLSDNIKILGPTPKPIARTHNLYHYQIILKYRFEDNLEETLNRILDWSQEADNRH 790

Query: 784 LRLAIDHEPQNM 796
 L+L ID EPQ +
 Sbjct: 791 LKLIIDCEPQQFL 803

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 vaccines or diagnostics.

Example 984

A DNA sequence (GBSx1044) was identified in *S.galactiae* <SEQ ID 3011> which encodes the amino acid sequence <SEQ ID 3012>. This protein is predicted to be methionyl-tRNA formyltransferase (fmt). Analysis of this protein sequence reveals the following:

Possible site: 13
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1329(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB13446 GB:Z99112 methionyl-tRNA formyltransferase [Bacillus subtilis]
 Identities = 155/314 (49%), Positives = 221/314 (70%), Gaps = 7/314 (2%)

Query: 1 MTKLLEFMGTPDFSATVLKGILADGKYDVLAVVTQPDRAVGRKKEIKMTPVKEVALENNIP 60
 MT+++FMGTPDFS VL+ ++ DG Y+V+ VVTQPD RGRK + PVKE AL + IP
 Sbjct: 1 MTRIVFMGTPDFSVPLRTLIEDG-YEVVGVTQPD RPKGRKKVLT PPPVKEEALRHGIP 59

Query: 61 VYQPEKLSGSPLEQLMTLGADGIVTAAFGQFLPTKLLESVGFA-INVHASLLPKYRGGA 119
 V QPEK+ + E+E+++ L D IVTAAFGQ LP +LL+S + INVHASLLP+ RGGA
 Sbjct: 60 VLQPEKVRLEETIEKVLALKPDLIVTAAFGQILPKELLDSPKYGCINVHASLLPELRGA 119

Query: 120 PIHYAIINGEKEAGVTTIMEMVAKMDAGDMVSKASVEITDEDNVGTMFDR LAVVGRDLLLD 179
 PIHY+I+ G+K+ G+TIM MV K+DAGDM+SK V+I + DNVGT+ D+L+V G LL +

-1085-

- Sbjct: 120 PIHYSILQGKKKTGITIMYMVEKLDAGDMISKVEVDIEETDNVGTLHDKLSVAGAKLLSE 179
- Query: 180 TLPGYLSGDIKPIPNQNEEVSFSPNISPDDEERIDWNKSSRDIFNHVRGMPWPVVAHTLLE 239
T+P ++G I P Q+EE+ +++PNI ++E +DW+++ +++N +RG+ PWPVA+T L
- 5 Sbjct: 180 TVPNVIAGSISPEKQDEEKATYAPNIKREQELLDWSRTGEELYNQIRGLNPWPVAYTTLN 239
- Query: 240 GNRFKLY--EVTMSEGKSGSPGQVIAKTKNSLT VATG-DGAIELKSVQPAKPRMDIKDFL 296
G K++ + + PG V+A K + VATG + A+ L +QPAGK RM +DF+
- 10 Sbjct: 240 GQNLKIWASKKIAAPT TAEPGTVVAVEKEGII VATGNETALLLTELQPAKPKRMKGEDFV 299
- Query: 297 NGVGRNLEIGDKFG 310
G ++E GD G
- Sbjct: 300 RGA--HVEAGDVLG 311
- 15 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3013> which encodes the amino acid sequence <SEQ ID 3014>. Analysis of this protein sequence reveals the following:
- Possible site: 15
>>> Seems to have no N-terminal signal sequence
- 20 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.0730(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
- 25 An alignment of the GAS and GBS proteins is shown below.
- Identities = 217/310 (70%), Positives = 266/310 (85%)
- Query: 1 MTKLLFMGTPDFSATVLKGLADGKYDVLAVVTQPDRAVGRKKEIKMTPVKEVALENNIP 60
M KLEFMGTP FSATVLKG+L + Y++L VVTQPDRAVGRKK+IK+TPVK++ALE+ I
- 30 Sbjct: 1 MIKLLFMGTPQFSATVLKGLLDNPAYEILGVVTQPDRAVGRKKDIKVTPVKQLALEHGIS 60
- Query: 61 VYQPEKLSGSGPELEQLMTLGADGIVTAAGQFLPTKLLESVGFAINVHASLLPKYRGGAP 120
+YQPEKLSGS EL ++M LGADGI+TAAGQFLPT LL+SV FAINVHASLLPKYRGGAP
- 35 Sbjct: 61 IYQPEKLSGSQELIEIMGLGADGIITAAFGQFLPTILLDSVSFAINVHASLLPKYRGGAP 120
- Query: 121 IHYAIINGEKEAGVTIMEMVAKMDAGDMVSKASVEITDEDNVGTMFDRLAVVGRDLLLD 180
IHYAI+NG+KEAGVTIMEM+ +MDAGDMV+KAS I + DNVGT+F++LA++GRDLLLD+
- Sbjct: 121 IHYAIMNGDKEAGVTIMEMIKEMDAGDMVAKASTPILETDNVGTLFKELAIIGRDLLLD 180
- 40 Query: 181 LPGYLSGDIKPIPNQNEEVSFSPNISPDDEERIDWNKSSRDIFNHVRGMPWPVVAHTLLEG 240
LP YLSG++KPIQ+ + +FSPNIS+ E++DW S++++FNH+RGM PWPVAHT LEG
- Sbjct: 181 LPAYLSGELKPIPDHSGATFSPNISPEHEKLDWTMSNQEVFNHIRMNPNPVAHTFLEG 240
- 45 Query: 241 NRFKLYEVTMSEGKSGSPGQVIAKTKNSLT VATGDGAIELKSVQPAKPRMDIKDFLNGVG 300
R K+YE ++EG+G PGQV+ KTK SL +ATG GA+ L VQPAKGP+M I DFLNG+G
- Sbjct: 241 QRLKIYEAQLAEGEGLPGQVVVTKKSLVIATGQGALSLIVVQPAKPKMSIIDFLNGIG 300
- Query: 301 RNLEIGDKFG 310
R LE+GD G
- 50 Sbjct: 301 RKLEVGDIIIG 310

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 985

- 55 A DNA sequence (GBSx1045) was identified in *S.agalactiae* <SEQ ID 3015> which encodes the amino acid sequence <SEQ ID 3016>. This protein is predicted to be sunL protein (sun). Analysis of this protein sequence reveals the following:

Possible site: 59
>>> Seems to have no N-terminal signal sequence

-1086-

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.1677(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

10 >GP:CAA10711 GB:AJ132604 sunL protein [Lactococcus lactis]
 Identities = 222/434 (51%), Positives = 305/434 (70%), Gaps = 15/434 (3%)
 Query: 7 KSARGLALMTLEEVFDKGAYSNIALNKSLSKSRSDKRALVTEIVYGTVARKITLEWYL 66
 K+AR AL L ++F AY+NI+L+++L+ S LS D+ VT +VYG V++K LEWY+
 Sbjct: 3 KNARQTALDVLNDFGNDAYANISLDRNLRDSELSTVDKGFVTALVYGVVSKKALLEWYI 62
 Query: 67 SHFIVDRDKLELVVYHLLLLSLYQLLYLDNIPDHAIVNDAVTIAKNRGNKKGAEKLINAV 126
 + + K W LLLL++YQ+L++D +P A V++AV IAK R + + INAV
 Sbjct: 63 TPLLKKEPKP--WAKMLLLLTYYQVLFMDKVPISAAVDEAVKIAK-RHDGQATANFINAV 119
 Query: 127 LRR-VSSETLPEIASIKRQNKRYSVAYSMPVWLKLLIDQYGETRALAIMESLFFERNKAS 185
 LR + SE E + K + YSMP L+ K++ Q+G R I+ESL + + S
 Sbjct: 120 LRNFMRSEHRNE-----EPKDWETKYSMPKLLLDKVMVRQFGGKRTGEILESLEKPSHVS 173
 Query: 186 LRVTDLSQKQTIKETLNVRDSHIAETALVADSGNFASSTFFQDGLITIQDESSQLVAPTL 245
 LR D + E R S + ETAL+ADSGNF+ T FQ G ITIQDE+SQLVAP L
 Sbjct: 174 LRKIDPTV----EIAGTRPSLLTETALIADSGNFSITTEFQTGRITIQDETSQVAPQL 228
 Query: 246 KVSGNDQVLDACSAPGGKTSHIASYLTGTAVTALDLYDHKLELVMENAKRIGLSDKIKTK 305
 ++ G ++VLDAC+ARGGK++H+A YLTGT +TALDLY+HKL+L+ +NA+R ++DKI T+
 Sbjct: 229 ELEGTEEVLDACAAPGGKSTHMAQYLTGTGHITALDLYEHKLDLINQNAQRQHVADKITTQ 288
 Query: 306 KLDASKAHEYFLEDTFDKILVDAPCSGIGLIRRPDIKYNKANQDFEALQEIQLSILSSV 365
 K DA+ +E F + FD+ILVDAPCSGIGLIRRPDI+Y K + DF LQ+IQL IL+S
 Sbjct: 289 KADATMIYENFGPEKFDRIILVDAPCSGIGLIRRPDIYRKESDFIDLQKIQLLEILNSA 348
 Query: 366 CQTLRKGGIITYSTCTIFEEENFQVIEKFLENHNPFEQVELSHTQEDIVKRGCSISPEQ 425
 ++L+K GI+ YSTCTIF+EEENF V+ +FLENHNPFEQVE+S+ + +++K GC+ I+PE
 Sbjct: 349 SKSLKSGIMVYSTCTIFDEENFVVEHFELENHNPFEQVEISNEKPEVIKEGCLFITPEM 408
 Query: 426 YHTDGGFFIGQVKRI 439
 YHTDGGFFI + K+I
 Sbjct: 409 YHTDGGFFIAKFKKI 422

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3017> which encodes the amino acid sequence <SEQ ID 3018>. Analysis of this protein sequence reveals the following:

45 Possible site: 42
 >>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

50 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

55 >GP:CAA10711 GB:AJ132604 sunL protein [Lactococcus lactis]
 Identities = 208/433 (48%), Positives = 287/433 (66%), Gaps = 13/433 (3%)
 Query: 7 KSTRGKALLVIEAIFDQAYTNIALNQQLSNKALSAKDRALLTEIVYGTVSRKISLEWYL 66
 K+ R AL V+ IF AY NI+L++ L + LS D+ +T +VYG VS+K LEWY+
 Sbjct: 3 KNARQTALDVLNDFGNDAYANISLDRNLRDSELSTVDKGFVTALVYGVVSKKALLEWYI 62
 Query: 67 AHYVKDRDKLDKVVYLLMLSLYQLTYLDKLEPAHAIVNDAVGIKAKNRGNKKGAEKFNVAI 126
 +K K W LL+L++YQ+ ++DK+P A V++AV IAK R + + P+NA+
 Sbjct: 63 TPLLKKEPKP--PWAKMLLLLTYYQVLFMDKVPISAAVDEAVKIAK-RHDGQATANFINAV 119

-1087-

Query: 127 LRQFTSHPLPDMETIKRRNKYYSVKYSLPVWLVKKLEDQFGSDRSVAIMESLFVRSKASI 186
 LR F E K + KYS+P L+ K+ QFG R+ I+ESL S S+
 Sbjct: 120 LRNFMR-----EHRNEEPKDWETKYSMPKLLDKMVRQFGGKRTGEILESLEKPSHVSL 174

5 Query: 187 RVTDPKLEEVAAEALDAERSLLSATGLTKASGHFAASDYFTNGDITIODESSQLVAPTLN 246
 R DP E SLL+ T L SG+F+ ++ F G ITIQDE+SQLVAP L
 Sbjct: 175 RKIDP-----TVEIAGTRPSLLTETALIADSGNF SITEEFQTGRITIQDETSQVLAPQLE 229

10 Query: 247 IDGDDIILDACSAPGGKTSHIASYLTKGVIALDLYDHKLELVKENANRLGVADNIETRK 306
 ++G + +LDAC+APGGK++H+A YL TG + ALDLY+HKL+L+ +NA R VAD I T+K
 Sbjct: 230 LEGTEEVLDAACAAPGGKSTHMAQVLTGHTALDLYEHKLDLQNAQRQHVADKITQK 289

15 Query: 307 LDAREVHRHFEKDSFDKILVDAPCSGIGLIRRKPDIKYNKESQGFNALQAIQLEILSSVC 366
 DA ++ +F + FD+ILVDAPCSGIGLIRRKPDY+Y KES F LQ IQLEIL+S
 Sbjct: 290 ADATMIYENFGPEKFDRIKVDAPCSGIGLIRRKPDIRYRKESDFIDLQKIQLEILNSAS 349

20 Query: 367 QTLRKGGIITYSTCTIFDEENRQVIRAFLOSHPNFEQVKLNHTQADIVKDGYLITPEQY 426
 ++L+K GI+ YSTCTIFDEEN V+ FL++HPNFEQV++++ + +++K+G L ITPE Y
 Sbjct: 350 KSLKSGIMVYSTCTIFDEENFDVHFELENHPNFEQVEISNEKPEVIKEGCLFITPEMY 409

25 Query: 427 QTDGFFIGQVRRV 439
 TDGFFI + +++
 Sbjct: 410 HTDGFFIAKFKKI 422

25 An alignment of the GAS and GBS proteins is shown below.

Identities = 305/440 (69%), Positives = 370/440 (83%)

Query: 1 MANDWKKSARGLALMTLEEVFDKGAYSNIALNKSLSKSRSDKDRALVTEIVYGTVARKI 60
 +A++WKKS RG AL+ +E +FD+GAY+NIALN+ L LS KDRAL+TEIVYGTV+RKI
 30 Sbjct: 1 LADNWKKSSTRGKALLVIEAIFDQAYTNIALNQQLSNKALSADKDRALLTEIVYGTVSRKI 60

Query: 61 TLEWYLSHFIVDRDKLELVVYHLLLSLYQLLYLDNIPDHAIVNDAVTIAKNRGNKKGAE 120
 +LEWYL+H++ DRDKL+ WVY+LL+LSLYQL YLD +P HAIVNDAV IAKNRGNKKGAE
 35 Sbjct: 61 SLEWYLAHVVKDRDKLDKWVYLLMLSLYQLTYLDKLPAAHAIVNDAVGIKNRGNKKGAE 120

Query: 121 KLINAVLRRVSSETLPEIASIKRQNKRYSVAYSMPVWLVKKLIDQYGETRALAIMESLFE 180
 K +NA+LR+ +S LP++ +IKR+NK YSV YS+PVWLVKKL DQ+G R++AIMESLF
 40 Sbjct: 121 KFNAILRQFTSHPLPDMETIKRRNKYYSVKYSLPVWLVKKLEDQFGSDRSVAIMESLFV 180

Query: 181 RNKASLRVTDLSQKQTIKETLNVRDSHIAETALVADSGNFASTSFFQDGLITIQDESSQL 240
 R+KAS+RVTD + + + E L+ S ++ T L SG+FA++ +F +G ITIQDESSQL
 45 Sbjct: 181 RSKASIRVTDPLKLEEVAAEALDAERSLLSATGLTKASGHFAASDYFTNGDITIODESSQL 240

Query: 241 VAPTLKVSGNDQVLDACSAPGGKTSHIASYLTTGAVTALDLYDHKLELV MENAKRLGLSD 300
 VAPTL + G+D +LDACSAPGGKTSHIASYL TG V ALDLYDHKLELV ENA RLG++D
 50 Sbjct: 241 VAPTLNIDGDDIILDACSAPGGKTSHIASYLTKGVIALDLYDHKLELVKENANRLGVAD 300

Query: 301 KIKTKKLDASKAHEYFLEDTFDKILVDAPCSGIGLIRRKPDIKYNKANQDFEALQEIQLS 360
 I+T+KLDA + H +F +D+FDKILVDAPCSGIGLIRRKPDIKYNK +Q F ALQ IQL
 55 Sbjct: 301 NIETRKLDAREVHRHFEKDSFDKILVDAPCSGIGLIRRKPDIKYNKESQGFNALQAIQLE 360

Query: 361 ILSSVCQTLRKGGIITYSTCTIFEEENFQVIEKFLNHPNFEQVELSHTQEDIVKRGCI 420
 ILSSVCQTLRKGGIITYSTCTIF+EEN QVIE FL++HPNFEQV+L+HTQ DIVK G +
 60 Sbjct: 361 ILSSVCQTLRKGGIITYSTCTIFDEENRQVIEAFLOSHPNFEQVKLNHTQADIVKDGyli 420

Query: 421 ISPEQYHTDGFFIGQVKRIL 440
 I+PEQY TDGFFIGQV+R+L
 Sbjct: 421 ITPEQYQTDGFFIGQVRRVL 440

60 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1088-

Example 986

A DNA sequence (GBSx1046) was identified in *S.agalactiae* <SEQ ID 3019> which encodes the amino acid sequence <SEQ ID 3020>. This protein is predicted to be pppL protein. Analysis of this protein sequence reveals the following:

```

5   Possible site: 45
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
10      bacterial cytoplasm --- Certainty=0.5796(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15   >GP:CAA10712 GB:AJ132604 pppL protein [Lactococcus lactis]
    Identities = 131/245 (53%), Positives = 177/245 (71%), Gaps = 4/245 (1%)

    Query: 1  MEISLLTDIGQRRSNNQDFINQFENKAGVPLIILADGMGGHRAGNIASEMVTDLGSDWA 60
              ME S+L+DIG +RS NQD++ + N+AG L +LADGMGGH+AGN+AS++TV DLG W+
    Sbjct: 1  MEYSILSDIGSKRSTNQDYVGTYVNRAGYQLFLLADGMGGHKAGNVASKLTVEDLGKLWS 60

20   Query: 61 ETDF---SELSEIRDWMLVSIETENRKIYELGQSDDYKGMGTTIEAVAIVGDNIIFAHVG 117
              ET F + + + W+ + EN I LG+ D+Y+GMGTT+EA+ I G+ I+ AHVG
    Sbjct: 61 ETFFDAGTPEATLEIWLNRQVRNENENIASLGKLDEYQGMGTTLEALVIKGNITIVSAHVG 120

25   Query: 118 DSRIGIVRQGEYHLLTSDHSLVNLVKAGQLTEEEAASHPQKNIIITQSIGQANPVEPDLG 177
              DSR ++R GE + +T+DHSLV ELV AGQ+TEEEA HP KNIIT+S+GQ N V+ D+
    Sbjct: 121 DSRTYLMRDGELNKIITTDHSLVQLVLDAGQITEEEAEVHPNKNIIITRSLGQTNEVQADIQ 180

30   Query: 178 VHLLIEEGDYLVVNSDGLTNMLSNADIAITVLITQEK-TLDDKNQDLITLANHRGGLDNITVA 236
              L+ GD +++NSDGLTNM+S +I VL +E TLD+K++ LI LAN GGLDNITV
    Sbjct: 181 ALELQAGDIILMNSDGLTNMVSTTEIMEVLEREDLTLDNKSEALIRLANEHGGLDNITVV 240

    Query: 237 LVYVE 241
              L+ E
35   Sbjct: 241 LIKFE 245

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3021> which encodes the amino acid sequence <SEQ ID 3022>. Analysis of this protein sequence reveals the following:

```

40   Possible site: 43
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
45      bacterial cytoplasm --- Certainty=0.5301(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

    Identities = 180/245 (73%), Positives = 220/245 (89%)

50   Query: 1  MEISLLTDIGQRRSNNQDFINQFENKAGVPLIILADGMGGHRAGNIASEMVTDLGSDWA 60
              M+ISL TDIGQ+RSNNQDFIN+F+NK G+ L+ILADGMGGHRAGNIASEMVTDLG +W
    Sbjct: 1  MKISLKTLDIGQKRSNNQDFINKFDNKKGITLVILADGMGGHRAGNIASEMVTDLGREWV 60

55   Query: 61 ETDFSELSEIRDWMLVSIETENRKIYELGQSDDYKGMGTTIEAVAIVGDNIIFAHVGDSR 120
              +TDF+ELS+IRDW+ +I++EN++IY+LGQS+D+KGMGTT+EAVA+V + I+AH+GDSR
    Sbjct: 61 KTDFTELSQIRDWLFETIQSENQRIYDLGQSEDFKGMGTTVEAVALVESSAIYAHIGDSR 120

    Query: 121 IGIVRQGEYHLLTSDHSLVNLVKAGQLTEEEAASHPQKNIIITQSIGQANPVEPDLGVLH 180
              IG+V G Y LLTSDHSLVNLVKAGQ+TEEEAASHPQ+NIITQSIGQA+PVEPDLG V +
60   Sbjct: 121 IGLVHDGHYTLTLLTSDHSLVNLVKAGQITEEEAASHPQRNIITQSIGQASPVPEPDLGVRV 180

```

-1089-

Query: 181 LEEGDYLVVNSDGLTNMLSNADIATVLTQEKTLDDKNQDLITLANHRGGLDNITVALVYV 240
 LE GDYLV+NSDGLTNM+SN +I T+L + +LD+KNQ++I LAN RGGLDNIT+ALV+
 Sbjct: 181 LEPGDYLVVNSDGLTNMISNDEIVTILGSKVSLDEKNQEMIDLANLRGGLDNITIALVHN 240

5 Query: 241 ESEAV 245
 ESE V
 Sbjct: 241 ESEDV 245

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 10 vaccines or diagnostics.

Example 987

A DNA sequence (GBSx1047) was identified in *S.galactiae* <SEQ ID 3023> which encodes the amino acid sequence <SEQ ID 3024>. Analysis of this protein sequence reveals the following:

15 Possible site: 56
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -10.03 Transmembrane 346 - 362 (340 - 372)

----- Final Results -----
 20 bacterial membrane --- Certainty=0.5012(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9539> which encodes amino acid sequence <SEQ ID 9540> was also identified.

25 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA10713 GB:AJ132604 hypothetical protein [Lactococcus lactis]
 Identities = 219/380 (57%), Positives = 284/380 (74%), Gaps = 8/380 (2%)

30 Query: 1 MIQIGKLFAGRYRILKSIGRGGMADVYLARDLILDNEEVAIKVLRITNYQTDQIAVARFQR 60
 MIQIGK+FA RYRI+K IGRGGMA+VY D L + +VAIKVLR+N++ D IA+ARFQR
 Sbjct: 1 MIQIGKIFADRYRIIKEIGRGGMANVYQGEDTFLGDRKVAIKVLRNPFENDDIAIARFQR 60

Query: 61 EARAMAELTHPNIVAIRDIGEEDGQQFLVMEYVDGFDLKKYIQDNAPLSNNEVVRIMNEV 120
 EA AMAEL+HPNIV I D+GE + QQ++VME+VDG LK+YI NAPL+N+E + I+ E+
 35 Sbjct: 61 EAFAMAELSHPNIVGISDVGEFESQQYIVMEFVDGMTLKQYINQNAPLANDEAIEIITEI 120

Query: 121 LSAMSLAHQKGIVHRDLKPQNILLTKKGTVKVTDGFI AVAFETSLTQTNSMLGSVHYLS 180
 LSAM +AH GI+HRDLKPQN+L+++ GTVKVTDGFI A +ETSLTQTN+M GSVHYLS
 40 Sbjct: 121 LSAMDMAHSHGIIHRDLKPQNVLVSSSGTVKVTDFGIAKALSETSLTQTNTMFGSVHYLS 180

Query: 181 PEQARGSKATVQSDIYAMGIMLFEMLTGHIPYDGS AVTIALQHFQKPLPSILAENKSV 240
 PEQARGS ATVQSDIYA+GI+LFE+LTG IP+DGDSAV IAL+HFQ+ +PSI+ N VP
 Sbjct: 181 PEQARGSNATVQSDIYAIGIILFELLTGQIPFDGDSAVAIALKHFQENIPSIINLNPEVP 240

45 Query: 241 QALENIVIKATAKLTDRYKTTYEMGRDLSTALSSTRHREPKL VFN-DTESTKTLPKVTS 299
 QALEN+VIKATAK + +RY EM D++T+ S R E KLVFN D + TK +P +
 Sbjct: 241 QALENVVIKATAKDINNRYADVEEMMTDVATSTSLDRRGEEKLVFNKDHDETKIMP--AN 298

Query: 300 TVSSLTTEQLLRNQKQAKTTEKITPDSASNDKTKSKKKASHRLG TIMKLF FALCVVGII 359
 ++ T+ L+ K+ EK +S++ + K+K K S + G I+ L L V+G
 50 Sbjct: 299 LINPYDTKPLI--DKKTDQEK AQSESSTTENNKNKNKSKK--GLIISLVVLLLVIGGG 354

Query: 360 VFAYKILVSPPTIRVPDVSN 379
 FA+ + +PT ++VP+V+N
 55 Sbjct: 355 AFAWAV-STPTNVKVPNVIN 373

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3025> which encodes the amino acid sequence <SEQ ID 3026>. Analysis of this protein sequence reveals the following:

-1090-

Possible site: 56
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -8.60 Transmembrane 349 - 365 (340 - 370)

5 ----- Final Results -----
 bacterial membrane --- Certainty=0.4439(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

10 The protein has homology with the following sequences in the databases:

>GP:CAA10713 GB:AJ132604 hypothetical protein [Lactococcus lactis]
 Identities = 209/378 (55%), Positives = 273/378 (71%), Gaps = 8/378 (2%)

15 Query: 1 MIQIGKLFAGRYRILKSIGRGGMADVYLANDLILDNEDVAIKVLRITNYQTDQVAVARFQR 60
 MIQIGK+FA RYRI+K IGRGGMA+VY D L + VAIKVLN++ D +A+ARFQR
 Sbjct: 1 MIQIGKIFADRYRIIKEIGRGGMANVYQGEDTFLGDRKVAIKVLRNPFENDDIAIARFQR 60

20 Query: 61 EARAMAELNHPNIVAIRDIGEEDGQQFLVMEYVDGADLKRYIQNHAPLSNNEVVRIMEEV 120
 EA AMAEL+HPNIV I D+GE + QQ++VME+VDG LK+YI +APL+N+E + I+ E+
 Sbjct: 61 EAFAMAELSHPNIVGISDVGEFESQQYIVMEFVDGMTLKQYINQNAPLANDEAIEIITEI 120

25 Query: 121 LSAMTLAHQKGIVHRDLKPQNILLTKEGVVKVTDGFIQVAFQETSLTQINSMLGSHVHLS 180
 LSAM +AH GI+HRDLKPQN+L++ G VKVTDGFIQ A +ETSLTQTN+M GSVHLS
 Sbjct: 121 LSAMDMAHSHGIIHRDLKPQNVLVSSSGTVKVTDFGIQAKALSETSLTQINTMFGSVHLS 180

30 Query: 181 PEQARGSKATIQSDIYAMGIMLFEMLTGHIPIYDGDSAVTIALQHFQKPLPSIIENHNVP 240
 PEQARGS AT+QSDIYA+GI+LFE+LTG IP+DGDSAV IAL+HFQ+ +PSII N VP
 Sbjct: 181 PEQARGSNATVQSDIYAIGIILFELLTGQIPFDGDSAVAIALKHQFQENIPSIINLNPEVP 240

35 Query: 241 QALENVVIRATAKKLSRDYGSTFEMSRDLMTALSYNRSRERKIIF-ENVESTKPLPKVAS 299
 QALENVVI+ATAK +++RY EM D+ T+ S +R E K++F ++ + TK +P
 Sbjct: 241 QALENVVIKATAKDINNRYADVEEMMTDVATSTSLDRRGEKLVFNKDHDETKIMPANLI 300

40 Query: 300 GPTASVSKLSPPTPTVLTQESRLDQNTQDALQPPTKKKKSGRFLGTLFKILFSFFIVGVA 359
 P + L QE +++ T+ + KK K G + + +L ++G
 Sbjct: 301 NPYDTKPLIDKKTD--DQEKQAQSESSTENNKNKNKKSKKGLIISLVVLLL---VIGGG 354

Query: 360 LFTYLILTKPTSVKVPNV 377
 F + + T PT+VKVPNV
 Sbjct: 355 AFAWAVST-PTNVKVPNV 371

An alignment of the GAS and GBS proteins is shown below.

Identities = 390/643 (60%), Positives = 480/643 (73%), Gaps = 29/643 (4%)

45 Query: 1 MIQIGKLFAGRYRILKSIGRGGMADVYLARDLILDNEEVAIKVLRITNYQTDQIAVARFQR 60
 MIQIGKLFAGRYRILKSIGRGGMADVYLA DLILDNE+VAIKVLRITNYQTDQ+AVARFQR
 Sbjct: 1 MIQIGKLFAGRYRILKSIGRGGMADVYLANDLILDNEDVAIKVLRITNYQTDQVAVARFQR 60

50 Query: 61 EARAMAELTHPNIVAIRDIGEEDGQQFLVMEYVDGFDLKKYIQDNAPLSNNEVVRIMNEV 120
 EARAMAEL HPNIVAIRDIGEEDGQQFLVMEYVDG DLK+YIQ++APLSNNEVVRIM EV
 Sbjct: 61 EARAMAELNHPNIVAIRDIGEEDGQQFLVMEYVDGADLKRYIQNHAPLSNNEVVRIMEEV 120

55 Query: 121 LSAMSLAHQKGIVHRDLKPQNILLTKKGTVKVTDGFIQVAFQETSLTQINSMLGSHVHLS 180
 LSAM+LAHQKGIVHRDLKPQNILLTK+G VKVTDGFIQVAFQETSLTQINSMLGSHVHLS
 Sbjct: 121 LSAMTLAHQKGIVHRDLKPQNILLTKEGVVKVTDGFIQVAFQETSLTQINSMLGSHVHLS 180

60 Query: 181 PEQARGSKATVQSDIYAMGIMLFEMLTGHIPIYDGDSAVTIALQHFQKPLPSILAENKSV 240
 PEQARGSKAT+QSDIYAMGIMLFEMLTGHIPIYDGDSAVTIALQHFQKPLPSI+ EN +VP
 Sbjct: 181 PEQARGSKATIQSDIYAMGIMLFEMLTGHIPIYDGDSAVTIALQHFQKPLPSIIENHNVP 240

65 Query: 241 QALENIVIKATAKKLTDYRKITYEMGRDLSTALSSTRHREPKLVFNDETESTKTLPKVTS- 299
 QALEN+VI+ATAKKL+DRY +T+EM RDL TALS R RE K++F + ESTK LPKV S
 Sbjct: 241 QALENVVIRATAKKLSRDYGSTFEMSRDLMTALSYNRSRERKIIFENVESTKPLPKVASG 300

Query: 300 -----TVSSLTTEQLLRNQAKTTEKITPDSASNDKTKSKKKASHRLLGITIMKL 349
 T + LT E L Q T+ + P + KKK S R LGT+ K+

-1091-

Sbjct: 301 PTASVKLSPPPTPTVLTLQESRL---DQTNQTDALQPPT-----KKKKSGRFLGTLFKI 349

Query: 350 FFALCVVGLIIVFAYKILVSPPTTIRVPDVSNKTVQAQKMTLENSGLKVGAI RNIESDSVSE 409
 F+ +VG+ +F Y IL PT+++VP+V+ ++ AK L + GLKVG IR IESD+V+E

5 Sbjct: 350 LFSFFIVGVALFTYLILTKPTSVKVPNVAGTSLKVAQQLYDVGLKVGKIRQIESDTVAE 409

Query: 410 GLVVKTDPAAGRSRREGAKVNLVIATPNKSFTLGNYKEHNYKDILKDL-QGKGVKKSLIK 468
 G VV+TDP AG ++R+G+ + LY++ NK F + NYK +Y++ + L + GV KS IK

10 Sbjct: 410 GNVVRTDPKAGTAKRQGSSTILYVSIGNKGFDMENYKGLDYQEA MNSLIETYGVPSKIK 469

Query: 469 VKRKINNDYTTGTILAQSLPEGTSFNPDPGNKKLTLTVAVNDPMIMPDTVGMTVGEVIETL 528
 ++R + N+Y T+++QS G FNP+G K+TL+VAV+D + MP VT + + + TL

15 Sbjct: 470 IERIVTNEYPTNTVISQSPSAGDKFNPNGKSKITLSVAVSDTITMPMVTEYSYADAVNTL 529

Query: 529 TDLGLDADNLVFIQMQNGV---YQTVVTPPSSSKIASQDPYGGVGLRRGDKVKLYLLG 585
 T LG+DA + Y + + + +P S + ++ Q PYYG + L ++ LYL

20 Sbjct: 530 TALGIDASRIKAYVPSSSSATGFVPIHSPSSKAIVSGQSPYYGTSLSLSDKGEISLYLYP 589

Query: 586 SKTTNNSSSTPIDSSASSSTGTTTSDSVSSSTDASTSDSSSTS 628
 +T ++SSS+ SS SSS ++ +DS + ++ S S +TS

Sbjct: 590 BETHSSSSSS---SSTSSSNSSSINDSTAPGSNTELPSETTS 629

SEQ ID 3024 (GBS297) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 43 (lane 6; MW 75kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 27 (lane 4; MW 100.2kDa) and in Figure 159 (lane 2-4; MW 100kDa). GBS297-GST was purified as shown in Figure 223, lane 3. GBS297-His was purified as shown in Figure 203, lane 8.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

30 Example 988

A DNA sequence (GBSx1048) was identified in *S.galactiae* <SEQ ID 3027> which encodes the amino acid sequence <SEQ ID 3028>. Analysis of this protein sequence reveals the following:

Possible site: 16

>>> Seems to have an uncleavable N-term signal seq

35 INTEGRAL Likelihood = -7.91 Transmembrane 60 - 76 (50 - 90)
 INTEGRAL Likelihood = -7.43 Transmembrane 7 - 23 (3 - 25)
 INTEGRAL Likelihood = -5.68 Transmembrane 27 - 43 (24 - 46)

----- Final Results -----

40 bacterial membrane --- Certainty=0.4163(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

45 >GP:BAB03323 GB:AB035448 hypothetical protein [Staphylococcus aureus]
 Identities = 53/230 (23%), Positives = 104/230 (45%), Gaps = 14/230 (6%)

Query: 5 QFFLLVEAVVLVMGLMKILSDDWTSFIFILAL--ILLALRF-YNNSRHNFLLTSLLLL 61
 Q ++ A++++ I + F+ +L L +L+ + + Y + R LL+

50 Sbjct: 9 QMLIIFTALMIIANFYIIFBK-IGFLLVLLGCVLVVGVLYPHKIRGLLAFWIGALLI 67

Query: 62 FLIFMLNPY-IIAAVFAVLVLIHFSQVKKKNRYALIQFKNHQLDVKTTRNQWLGTDO 120
 + N Y II VF +L ++ + K K A + +K +W G +

55 Sbjct: 68 AFTLLSNKYTIIILFVFLLLLVRYLIHKFKPKKVATDEVMTSPSFIK---QKWFGEQR 124

Query: 121 HESDFYAFEDINIIRISGTD'TIDLTVNVIVSGQDNVIIIQKVFQDTRKVLVPLDVAVKADIS 180
 Y +ED+ I G IDLT ++N I+++ + G +V++P++ + ++

-1092-

Sbjct: 125 TPVYVYKWEDVQIQHGIGDLHIDLTKAANIKENNTIVVRHILGKVQVILPVNINLHVA 184

Query: 181 SVYGSVQYFDFFEEYDLRNESIKLSQ--EEYYLLKRVKLVVNTIAGKVEV 228

+ YGS Y + + Y + N + I + + + + Y V + V + T G VEV

5 Sbjct: 185 AFYGST-YVNEKSYKVENNNIHIEEMMKPDNY---TVNIYVSTFIGDVEV 230

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3029> which encodes the amino acid sequence <SEQ ID 3030>. Analysis of this protein sequence reveals the following:

10 Possible site: 35
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -9.92 Transmembrane 44 - 60 (36 - 64)
 INTEGRAL Likelihood = -8.76 Transmembrane 69 - 85 (66 - 105)
 INTEGRAL Likelihood = -8.70 Transmembrane 24 - 40 (20 - 42)
 15 INTEGRAL Likelihood = -6.64 Transmembrane 88 - 104 (85 - 105)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.4970(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 20 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:BAB03323 GB:AB035448 hypothetical protein [Staphylococcus aureus]
 Identities = 41/187 (21%), Positives = 85/187 (44%), Gaps = 22/187 (11%)
 25 Query: 47 FILILVL--ILLALRF-YNQDSRNNFLLTVSLFLFLIFMLNPYIIMAVLLGIVYIFINH 103
 F+L+L+L +L+ + + Y R + L + + N Y I+ + + + + + +
 Sbjct: 33 FLLVLLLGCVLVYVGILYFHKIRGLLAFWIGALLIAFTLLSNKYITIIILFVFLLLIV-- 90
 30 Query: 104 FSQVKKKNRFALIRFKEEKIEVNNT-----KHQWIGTANYESDYCFDDINIIRISG 155
 R+ + +FK +K+ + K+W G Y ++D+ I G
 Sbjct: 91 -----RYLIHKFKPKKVATDEVMTSPSFIKQKWFGEQRTPVYVYKWEDVQIQHGIG 142
 35 Query: 156 NDTVDLTNVIVTGMNDNIIVIRKIFGNTTILVPIDVTVTLDVSSIIYGSVDFFRCQQYDLRN 215
 + +DLT +N IV+R I G +++P++ + L V++ YGS + + Y + N
 Sbjct: 143 DLHIDLTKAANIKENNTIVVRHILGKVQVILPVNINLHVAAFYGST-YVNEKSYKVEN 201
 Query: 216 ESIKFKE 222
 +I +E
 40 Sbjct: 202 NNIHIEE 208

An alignment of the GAS and GBS proteins is shown below.

Identities = 137/211 (64%), Positives = 175/211 (82%)
 45 Query: 1 MKKFQFFLLVEAVVLVGLMKILSDDWTSFIFILALILLALRFYNNDNRHINFLLTSLLL 60
 MKKFQFFLL+E ++L MG+M IL +D +SFI IL LILLALRFYN DSR+NFLLT SLL
 Sbjct: 18 MKKFQFFLLIECILLAMGIMTILNDLSSFILILVLILLALRFYNQDSRNNFLLTVSLLF 77
 50 Query: 61 LFLIFMLNPYIIAAVVFVAVLYVLINHFSQVKKKNRYALIQFKNHQLDVKTTRNQWLGTDO 120
 LFLIFMLNPYII AV+ ++Y+ INHFSQVKKKNR+ALI+FK +++V T++QW+GT
 Sbjct: 78 LFLIFMLNPYIIMAVLLGIVYIIFINHFSQVKKKNRFALIRFKEEKIEVNNTKHQWIGTAN 137
 Query: 121 HESDFYAFEDINIIRISGTDITDLTNVIVSGQDNVIIQKVFQVGLVPLDVAVKADIS 180
 +ESD+Y F+DINIIRISG DT+DLTNVIV+G DN+I+I+K+FG+T +LVP+DV V D+S
 55 Sbjct: 138 YESDYCFDDINIIRISGNDTVDLTNVIVTGMNDNIIVIRKIFGNTTILVPIDVTVTLVDS 197
 Query: 181 SVYGSVQYFDFFEEYDLRNESIKLSQEEYYL 211
 S+YGSV +F ++YDLRNESIK + + L
 60 Sbjct: 198 SIYGSVDFFRCQQYDLRNESIKFKETDNQSL 228

SEQ ID 3028 (GBS66) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 6 (lane 4; MW 25kDa) and in Figure 7 (lane 2; MW 24.7kDa).

-1093-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 989

A DNA sequence (GBSx1049) was identified in *S.agalactiae* <SEQ ID 3031> which encodes the amino acid sequence <SEQ ID 3032>. This protein is predicted to be histidine kinase (narQ). Analysis of this protein sequence reveals the following:

```

Possible site: 19
>>> Seems to have an uncleavable N-term signal seq
10  INTEGRAL    Likelihood = -11.41    Transmembrane    47 - 63 ( 40 - 72)
    INTEGRAL    Likelihood = -9.98    Transmembrane     9 - 25 ( 5 - 36)

----- Final Results -----
        bacterial membrane --- Certainty=0.5564 (Affirmative) < succ>
        bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
15        bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:CAB54570 GB:AJ006393 histidine kinase [Streptococcus pneumoniae]
Identities = 159/334 (47%), Positives = 239/334 (70%), Gaps = 5/334 (1%)
20
Query: 1  MKKHHYFLAFFYGSVIFAI CFVIIDSLGVNL-VHLYQTSRLWLIEQLIFSIFFLSLAVT 59
      MKK  Y + . . . +F . . . L + + L+ + . E+ +F + S+++T
Sbjct: 1  MKKQAYVIIALTSFLFVFFFSHSLEILDFDWSIFLHDVEKT---EKFVFLLLVFSMSMT 57

25
Query: 60  ILLLLTWFLDDNSKRQINHNLRILNNQSINVTDDGTEISTNIQRLSKKMNLMTASLQS 119
      LL L W +++ S R++ NL+R+L Q + . D ++ + + LS K+NL+T +LQ
Sbjct: 58  CLLALFWRGIEELSLRKMQLNLRLLAGQEVVQVAD-PDLDA SFKSLSGKLNLLTEALQK 116

30
Query: 120 KENSRIKLSQEIVKQERKRIARDLHDTVSDQLFAASMVLSGIAQNVSQLDQVGSQLLA 179
      EN + + +EI+++ERKRIARDLHDTV SQ+LFAA M+LSGI+Q +LD +++ +QL +
Sbjct: 117 AENQSLAQEEETIEKERKRIARDLHDTV SQELFAAHMILSGISQALKLDREKMQTQLQS 176

35
Query: 180 VEEMLQHAQNDRILLHLRPVELENKTLSEGFRMILKELTDKSDIEVVYHESILTLPKK 239
      V +L+ AQ DLR+LLHLRPVELE K+L EG +++LKEK DKSD+ V +++ LPKK
Sbjct: 177 VTAILETAQKDLRVLLHLRPVELEQKSLIEGIQILLKELEDKSDLRVSLKQNTKLPKK 236

40
Query: 240 IEDNIFRIGQEFISNTLKHSQASRLLEVYLNQTENELQLK MIDNGIGFDMDSVYDLSYGLK 299
      IE++IFRI QE ISNTL+H+QAS L+VYL QT+ ELQLK++DNGIGF + S+ DLSYGL+
Sbjct: 237 IEEHIFRILQELISNTRLRHAQASCLDVLYQTDVELQLKVVDNGIGFQLGSLDDL SYGLR 296

Query: 300 NIEDRVEDLAGNLQLLSQPGKGVAMDIRLPLVNQ 333
      NI++RVED+AG +QLL+ P +G+A+DIR+PL+++
Sbjct: 297 NIKERVEDMAGTVQLLTAPKQGLAVDIRIPLLDK 330

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2991> which encodes the amino acid sequence <SEQ ID 2992>. Analysis of this protein sequence reveals the following:

```

Possible site: 18
>>> Seems to have an uncleavable N-term signal seq
50  INTEGRAL    Likelihood = -14.22    Transmembrane    49 - 65 ( 42 - 70)
    INTEGRAL    Likelihood = -6.58    Transmembrane     8 - 24 ( 5 - 33)

----- Final Results -----
        bacterial membrane --- Certainty=0.6689 (Affirmative) < succ>
        bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
55        bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

Identities = 218/337 (64%), Positives = 276/337 (81%), Gaps = 3/337 (0%)

-1095-

```

|::| || | :: | :: ||:| | : | : : | | :|| | : : |:::
SMSMTCLLALFWRGIEELSLRKMQANLKRLLAGQEVVQVAD-PDLDAFSLSGKLNLLTEALQKAENQSLAQEBEIEK
          70      80      90      100      110      120      130

5   732      762      792      822      852      882      912      942
   ERKRIARDLHDTVSQLFAASMLVSGIAQNVSQLDQVGSQQLLAVEEMLQHAQNDRILLLHLRPVELENKTLSEGFRL
   ||||| ||||| :||| |:|||: | || :|| :| :| :| |||:||||| ||| | :| || :|
   ERKRIARDLHDTVSQLFAAHMILSGISQQALKLDREKMQTQLQSVTAILETAQKDLRVLLHLRPVELEQKSLIEGIQI
          150      160      170      180      190      200      210

10  972      1002     1032     1062     1092     1122     1152     1182
   ILKELTDKSDIEVVYHESILTLPPKIEDNIFRIGQEFISNTLKHSQASRLLEVYLNQTENELQLKMIDNGIGFDMDSVYDL
   :||| | |||: | :: | ||||:| || | ||||:| || | ||| | ||| | |||:| ||||| : | : |
   LLKELEDKSDLRVSLKQNMTPKLPKIEEHI FRILQELISNLRHAQASCLDVLYQTDVELQLKVVDNGIGFQLGSLDDL
          230      240      250      260      270      280      290

15  1212     1242     1272     1302     1332     1362     1392     1422
   SYGLKNIXDRVEDLAGNLQLLSQPGKGVAMDIRLPLVNQSEDKNG*NKNCTC**P*DGSSRFKKFFKLTSC*SNR*GLK
   ||||:| :|||:| :|||: | :|:| |||:|:::
20  SYGLRNK KERVEDMAGTVQLLTAPKQGLAVDIRIPLLDKE
          310      320      330

```

SEQ ID 8702 (GBS31) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 15 (lane 8; MW 64kDa). It was also expressed as GBS31d in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 151 (lane 8-10; MW 59kDa) and in Figure 187 (lane 8; MW 59kDa). GBS31d was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 151 (lane 11-13; MW 34kDa) and in Figure 182 (lane 11; MW 34kDa). Purified GBS31d-GST is shown in lane 3 of Figure 237.

Example 990

A DNA sequence (GBSx1050) was identified in *S.agalactiae* <SEQ ID 3033> which encodes the amino acid sequence <SEQ ID 3034>. Analysis of this protein sequence reveals the following:

```

Possible site: 31
>>> Seems to have no N-terminal signal sequence

```

```

----- Final Results -----
          bacterial cytoplasm --- Certainty=0.2706(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:CAB54571 GB:AJ006393 response regulator [Streptococcus pneumoniae]
Identities = 154/209 (73%), Positives = 184/209 (87%)

```

```

Query: 8   IKIVLVDDHEMVRLGLKSFLNLQADVEVIGEASNGLEGIKKALELRPDVVVMDLVMPEDM 67
+KI+LVDDHEMVRLGLKS+ +LQ DVEV+GEASNG +GI ALELRPDV+VMD+VMPEDM+
Sbjct: 1   MKILVLDDHEMVRLGLKSYFDLQDDVEVVGESNGSQGIDLALRLPDVIVMDIVMPEDM 60

Query: 68  GVEATLALLKDWPEAAAILVLTSYLDNEKIYPVIEAGAKGYMLKTSSAAEILNAIRKVSRG 127
G++ATLA+LK+WPEA IL++TSYLDNEKI PV++AGAKGYMLKTSSA E+L+A+ KV+ G
Sbjct: 61  GIDATLAILKEWPEAKILIVTSYLDNEKIMPVLDAGAKGYMLKTSSADELLHAVSKVAAG 120

Query: 128 EQAIENEVDKKIKAHDKCPALHEGLTARERDILNLLAKGYDNQRIADELFISLKTIVKTHV 187
E AIE EV KK++ H LHE LTARERD+L L+AKGY+NQRIAD+LFISLKTIVKTHV
Sbjct: 121 ELAIEQEVSKKVEYHRNHMELHEELTARERDVLQLIAKGYENQRIADDLFISLKTIVKTHV 180

Query: 188 SNILGKLNVDRTQAVVYAFQHHLVPQDD 216
SNIL KL V+DRTQA VYAFQHHLV Q++
Sbjct: 181 SNILAKLEVS DRTQAAVYAFQHHLVGQEE 209

```

-1096-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2995> which encodes the amino acid sequence <SEQ ID 2996>. Analysis of this protein sequence reveals the following:

Possible site: 29

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3094(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 175/212 (82%), Positives = 192/212 (90%)

Query: 5 MDKIKIVLVDDHEMVRGLKSFNLQADVEVIGEASNGLEGIKKALELRPDVVVMDLVMP 64

M KIK++LVDDHEMVR+GLKSFNLQAD++V+GEASNG EG+ AL L+PDV+VMDLVMP

Sbjct: 3 MSKIKVILVDDHEMVRMGLKSFNLQADIDVVGEASNGREGVDLALALKPDVLVMDLVMP 62

Query: 65 EMDGVEATLALLKDWPEAAAILVLTSLDNEKIYPVIEAGAKGYMLKTSSAAEILNAIRKV 124

E+ GVEATL +LK W EA +LVLTSLDNEKIYPVI+AGAKGYMLKTSSAAEILNAIRKV

Sbjct: 63 ELGGVEATLEVLKKWKEAKVLVLTSLDNEKIYPVIDAGAKGYMLKTSSAAEILNAIRKV 122

Query: 125 SRGEQAIENEVDKKIKAHDKCPALHEGLTARERDILNLLAKGYDNQRIADELFISLKTVK 184

S+GE AIE EVDKKIKAH+ P LHE LTARE DIL+LLAKGYDNQ IADELFISLKTVK

Sbjct: 123 SKGELAIETEVDKKIKAH+QHPDLHEELTAREYDILHLLAKGYDNQTIADLFISLKTVK 182

Query: 185 THVSNILGKLNVDRTQAVVYAFQHHLVPQDD 216

THVSNIL KL V DRTQAVVYAF+HHLVPQDD

Sbjct: 183 THVSNILAKLEVGDRTQAVVYAFRHHLVPQDD 214

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 991

A DNA sequence (GBSx1051) was identified in *S.agalactiae* <SEQ ID 3035> which encodes the amino acid sequence <SEQ ID 3036>. Analysis of this protein sequence reveals the following:

Possible site: 59

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1688(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB08166 GB:Z94864 putative peptidyl-prolyl cis-trans isomerase

[Schizosaccharomyces pombe]

Identities = 81/174 (46%), Positives = 109/174 (62%), Gaps = 30/174 (17%)

Query: 288 IKTNHGDMTVKLFDPDHAPKTVANFIGLAKQGYDGIIFHRIIPDFMIQGGDPTGTGMGGE 347

++T+ G + ++L+ +HAPKT NF LAK+GYDGI+IFHR+IPDF+IQGGDPTGTG GG

Sbjct: 6 LQTSLGKILIELYTEHAPKTCQNFYTLAKEGYDGVIFHRVIPDFVIQGGDPTGTGRGGT 65

Query: 348 SIYGESFEDEFSEELYNV-RGALSMANAGPNTNGSQFFIVQNTKIPYAKKELERGGWPTP 406

SIYG+ F+DE +L++ G LSMANAGPNTN SQFFI T P

Sbjct: 66 SIYGDKFDEIHSDDLHHTGAGILSMANAGPNTNSSQFFI---TLAP----- 108

Query: 407 IAELYAGQGGTTPHLDRRHVSFVGLVDQSSFEVLDEIAAVETGSQDKPLEDVVIL 460

TP LD +H++FG++V S V + + T S D+P+E + I+

Sbjct: 109 -----TPWLDGKHITIFGRVV--SGLSVCKRMGLIRTDSSDRPIEPLKII 150

-1097-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3037> which encodes the amino acid sequence <SEQ ID 3038>. Analysis of this protein sequence reveals the following:

Possible site: 59

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2175(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 381/464 (82%), Positives = 422/464 (90%)

Query: 1 MDAKTKYKAKKIKAVFFDIDDTLRVKDTGYMPPSILKVFKALKDKGIVVGLIASGRARYGV 60

MDAK KYKAKKIK VFFDIDDTLRVKDTGYM SI +VFKALK KGI+VGLIASGRARYGV

Sbjct: 5 MDAKLKYKAKKIKMVFFDIDDTLRVKDTGYMPESIQRVFKALKAKGILVGLIASGRARYGV 64

Query: 61 PKEVQDLNADYCVKLNAYVKDKDKNIIFHRPIAEYVEQYKKWADTVGICYGLAGRHEA 120

P+EVQDL+ADYCVKLNAYVKD K IIF PIPA+ V YKKWAD +GI YG+AGRHEA

Sbjct: 65 PQEVQDLHADYCVKLNAYVKDDAKTIIFQAPIPADVVVAYKKWADDMGIFYGMAGRHEA 124

Query: 121 VLSRDDLNDVNDVIDVSDLEVPDFNKEHDIYQMWTTFEDKGDLSLHLEPELAEHLRLIRW 180

VLS R+D++++AID VY+ LEV PD+N+ HD+YQMWTTFEDKGD L LP LAEHLRL+RW

Sbjct: 125 VLSARNDMISNAIDNVYAQLEVCPCDYNEYHDVYQMWTTFEDKGDGLQLPAELAEHLRLVRW 184

Query: 181 HDHSSDVVLKGTSKALGVSKVVEHLGLKPENILVFGDELNDLELFDYAGLAVAMGVSHPE 240

HD+SSDVVLKGTSKALGVSKVV+HLGLKPENILVFGDELNDLELFDYAG+++AMGVSHP

Sbjct: 185 HDNSSDVVLKGTSKALGVSKVVDHLGLKPENILVFGDELNDLELFDYAGISIAMGVSHPL 244

Query: 241 AQKKADFITKKVEEDGILYALEELGLIEKELTFPQVDIENTEGPVAVIKTNHGDMTVKLF 300

Q+KADFITKKVEEDGILYALEELGLI+KEL FPQ+D+ N +GP A IKTNHGDMT+ LF

Sbjct: 245 LQEKADFITKKVEEDGILYALEELGLIDKELQFPQLDLPNHKGPKATIKTNHGDMTLVLF 304

Query: 301 PDHAPKTVANFIGLAKQGYDGIIFHRIIPDFMIQGGDPTGTGMGGESIYGESFEDEFSE 360

PDHAPKTVANF+GLAK+GYDGIIFHRIIP+FMQGGDPTGTGM G+SIYGESFEDEFSE+

Sbjct: 305 PDHAPKTVANFLGLAKEGYDGIIFHRIIPEFMIQGGDPTGTGMCGQSIYGESFEDEFSD 364

Query: 361 ELYNVRGALSMANAGPNTNGSQFFIVQNTKIPYAKKELERGGWPTPIAELYAGQGGTPhL 420

ELYN+RGALSMANAGPNTNGSQFFIVQN+KIPYAKKELERGGWP PIA YA GGTPhL

Sbjct: 365 ELYNLRGALSMANAGPNTNGSQFFIVQNSKIPYAKKELERGGWPAPIAASYAANGGTPhL 424

Query: 421 DRRHSVFGQLVDQSSFEVLDEIAAVETGSQDKPLEDVVILTIEV 464

DRRH+VFGQLVD++SF+VLD IA VETG+QDKP EDV+I TIEV

Sbjct: 425 DRRHTVFGQLVDETSFQVLDLIAGVETGAQDKPKEDVIIETIEV 468

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 992

A DNA sequence (GBSx1052) was identified in *S.agalactiae* <SEQ ID 3039> which encodes the amino acid sequence <SEQ ID 3040>. This protein is predicted to be ribosomal protein S1 (rpsA). Analysis of this protein sequence reveals the following:

Possible site: 36

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3126(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-1098-

The protein has homology with the following sequences in the GENPEPT database.

```

5  >GP:BA07066 GB:AP001518 polyribonucleotide nucleotidyltransferase
    (general stress protein 13) [Bacillus halodurans]
    Identities = 46/120 (38%), Positives = 71/120 (58%), Gaps = 11/120 (9%)

Query: 8  KIGDKLKGTVTGIRPYGAFVSLDGRTGLIHISEIKTGYIDNIYDVLVSGDEVYVQVIDV 67
      ++G  ++G VTGI+P+GAFV+++D + GL+HISE+  G++ +I DVLVSGDEV V+++ V
Sbjct: 5  EVGSIVEGKVTGIRKPFPGAFVAIDQKQGLVHISEVAHGFKVDINDVLVSGDEVKVKILSV 64

10 Query: 68  DEFTQKASLSLRTLEERHHIQH-----RHRFSNNRLKIGFKPLEENLPSWVEE 116
      DE + K SLS+R  +E                      R                      GF LE+ I  W+++
Sbjct: 65  DEESGKISLSIRATQEAPERPARAPKPRPAGGGGRKPQKQSQGQGFNTLEDKLEWLKQ 124

```

15 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3041> which encodes the amino acid sequence <SEQ ID 3042>. Analysis of this protein sequence reveals the following:

```

Possible site: 58
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
20  bacterial cytoplasm --- Certainty=0.1832(Affirmative) < succ>
    bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

25  Identities = 78/115 (67%), Positives = 100/115 (86%)

Query: 7  MKIGDKLKGTVTGIRPYGAFVSLDGRTGLIHISEIKTGYIDNIYDVLVSGDEVYVQVID 66
      MKIGDKL GT+TGI+PYGAFV+LE+G TGLIHISEIKTG+ID+I  +L++G++V VQVID
30  Sbjct: 1  MKIGDKLHGTITGIRPYGAFVLENGTGLIHISEIKTGFIDDDQLAIGNQVLVQVID 60

Query: 67  VDEFTQKASLSLRTLEERHHIQHRRHRFSNNRLKIGFKPLEENLPSWVEEGLAYL 121
      +DE+++K SLS+RTL EE+ H  HRHR+SN+R KIGF+PLEE LP W+EE L +L
Sbjct: 61  IDEYSKKPSLSMRTLAEKQHFHRRHRYNSNRHKIGFRPLEEQLPQWIEESLQFL 115

```

35 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 993

40 A DNA sequence (GBSx1053) was identified in *S.agalactiae* <SEQ ID 3043> which encodes the amino acid sequence <SEQ ID 3044>. This protein is predicted to be pyruvate formate-lyase 2 activating enzyme (pflA). Analysis of this protein sequence reveals the following:

```

Possible site: 41
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
45  bacterial cytoplasm --- Certainty=0.2889(Affirmative) < succ>
    bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

50  >GP:AAC76934 GB:AE000469 probable pyruvate formate lyase activating
    enzyme 2 [Escherichia coli K12]
    Identities = 90/251 (35%), Positives = 142/251 (55%), Gaps = 16/251 (6%)

Query: 8  VFNIQHFSIHGDPGIRTTVFLKGCPLRCPWCANPESQKMVPETMR----- 52
      +FNIQ +S++DG GIRT VF KGCP CPWCANPES  +T+R
55  Sbjct: 24  IFNIQRYSLNDGEGIRTVVFFKGCPLRCPWCANPESISGKIQTVRREAKCLHCAKCLRDA 83

```

-1099-

Query: 53 -DAITNESVIVGEEKSVDDIIEEVLKDIDFYEESSGGGITLSGGEIFAQFEFAKAILKRAK 111
 + + +G + S+D + EV+KD F+ SGGG+ILSGGE+ Q EFA L+R +
 Sbjct: 84 DECPSGAFERIGRDISLDALEREVMKDDIFFRTSGGGVTLSGGEVLMQAEFATRLQLRLR 143

5 Query: 112 SLGIHTAIETTAYTRHEQFIDLIQYVDFIYTDLKHYNLSLKHQEKTMVKNASIIKNIHYAF 171
 G+ AIET + + L + D + DLK ++ + ++ + +++N+
 Sbjct: 144 LWGVSCAIETAGDAPASKLLPLAKLCDEVLFDLKIMDATQARDVVKMNLPRVLENLRLLV 203

10 Query: 172 ANGKTIVLRIPVIPNFNDSLEDAEEFACLFDRDLIRQVQLLPFHQFGQNKYQLLNROYEM 231
 + G ++ R+P+IP F S E+ ++ + L+IRQ+ LLPFHQ+G+ KY+LL + + M
 Sbjct: 204 SEGVNVIPRLPLIPGFTLSRENMQQALDVLIPLNIRQIHLLPFHQYGEKPYRLLGKTWSM 263

Query: 232 EEIAALHPEDL 242
 +E+ A D+
 15 Sbjct: 264 KEVPAPSSADV 274

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3045> which encodes the amino acid sequence <SEQ ID 3046>. Analysis of this protein sequence reveals the following:

20 Possible site: 58
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2209(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 25 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 187/255 (73%), Positives = 220/255 (85%)

30 Query: 4 EKGIVFNIQHFSIHDGPGIRTTVFLKGCPLRCPWCANPESQKMPETMRDAITNESVIVG 63
 ++GIVFNIQHFSIHDGPGIRTTVFLKGCPLRCPWCANPESQ+ PE M + + IVG
 Sbjct: 3 DRGIVFNIQHFSIHDGPGIRTTVFLKGCPLRCPWCANPESQQKAPEQMLTSDGLNTRKIVG 62

35 Query: 64 EEKSVDDIIEEVLKDIDFYEESSGGGITLSGGEIFAQFEFAKAILKRAKSLGIHTAIETTA 123
 EEK+VD++IEEVLKD+DFYEESSGGG+ILSGGEIFAQF+FA A+LK AK+ G+HTAIETTA
 Sbjct: 63 EEXTVDEVIEEVLKDLDFYEESSGGMTLSGGEIFAQFDFALALLKAAKAAGLHTAIETTA 122

40 Query: 124 YTRHEQFIDLIQYVDFIYTDLKHYNLSLKHQEKTMVKNASIIKNIHYAFANGKTIVLRIPV 183
 + +HEQF+ L+ YVDFIYTDLKHYN L+HQ+ T V+N IIKNIHYAF GK IVLRIPV
 Sbjct: 123 FAKHEQFVTLVDYVDFIYTDLKHYNQLRHQKVTGVRNDLIIKNIHYAFQAGKEIVLRIPV 182

45 Query: 184 IPNFNDSLEDAEEFACLFDRDLIRQVQLLPFHQFGQNKYQLLNROYEMEEIAALHPEDLL 243
 IP FNDL+DA+ F+ LF++L+I QVQLLPFHQFG+NKY+LL R+YEM E+ A HPEDL
 Sbjct: 183 IPQFNDSLDDAKAFSELFNQLEIDQVQLLPFHQFGENKYKLLGREYEMAEVKAYHPEDLA 242

Query: 244 DYQAIFSKYNIHCYF 258
 DYQA+F +NIHCYF
 Sbjct: 243 DYQAVFLNHNHCYF 257

50 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 994

A DNA sequence (GBSx1054) was identified in *S.agalactiae* <SEQ ID 3047> which encodes the amino acid sequence <SEQ ID 3048>. Analysis of this protein sequence reveals the following:

55 Possible site: 57
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1762(Affirmative) < succ>
 60 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

-1100-

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9299> which encodes amino acid sequence <SEQ ID 9300> was also identified.

- 5 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC74366 GB:AE000226 putative DEOR-type transcriptional
regulator [Escherichia coli K12]
Identities = 74/177 (41%), Positives = 113/177 (63%), Gaps = 1/177 (0%)

10 Query: 2 N R L E N I I S L V S Q Y Q K I D V N T L S E L L Q V S K V T I R K D L D K L E G K G L L H R E H G Y A V L N S G D D L 61
+R + I+ +V ++ V L++ VS+VTIR+DL+ LE L R HG+AV DD+
Sbjct: 3 S R Q Q T I L Q M V I D Q G Q V S V T D L A K A T G V S E V T I R Q D L N T L E K L S Y L R R A H G F A V S L D S D D V 62

15 Query: 62 N V R L S F N H K T K K E I A A L A A N M V S D N D T I L I E S G S T C A L L A E N I C Q T K R N V T I L T N S C F I A 121
R+ N+ K+E+A AA++V +TI IE+GS+ ALLA + + K+NVTI+T S +IA
Sbjct: 63 E T R M S N Y T L K R E L A E F A A S L V Q P G E T I F I E N G S S N A L L A R T L G E Q K K N V T I I T V S S Y I A 122

Query: 122 N Y L R E Y D S C Q I V L L G G E Y Q S S S Q V T V G P L L K K M I S L F H V S L A F V G T D G F D P K T R I Y G 178
+ L++ C+++LLGG YQ S+ VGPL ++ I H S AF+G DG+ P+T G

20 Sbjct: 123 H L L K D - A P C E V I L L G G V Y Q K S E S M V G P L T R Q C I Q V H F S K A F I G I D G W Q P E T G F T G 178

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3049> which encodes the amino acid sequence <SEQ ID 3050>. Analysis of this protein sequence reveals the following:

25 Possible site: 58
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2888(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
30 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 131/171 (76%), Positives = 150/171 (87%)

35 Query: 1 M N R L E N I I S L V S Q Y Q K I D V N T L S E L L Q V S K V T I R K D L D K L E G K G L L H R E H G Y A V L N S G D D 60
M N R L E I I L V S Q + K I D V N + L S E L V S K V T I R K D L D K L E K G L L R E H G Y A V L N S G D D
Sbjct: 2 M N R L E I I Q L V S Q K K I D V N S L S E Q L D V S K V T I R K D L D K L E S K G L L R R E H G Y A V L N S G D D 61

40 Query: 61 L N V R L S F N H K T K K E I A A L A A N M V S D N D T I L I E S G S T C A L L A E N I C Q T K R N V T I L T N S C F I A 120
L N V R L S + N + K + I A A A + V D N D T I + I E S G S T C A L L A E + C Q T K R N + + + T N S C F I
Sbjct: 62 L N V R L S Y N Y N I K R R I A E K A A E L V Q D N D T I M I E S G S T C A L L A E V L C Q T K R N I K V I T N S C F I 121

Query: 121 A N Y L R E Y D S C Q I V L L G G E Y Q S S S Q V T V G P L L K K M I S L F H V S L A F V G T D G F D 171
A N Y + R + Y S C Q I + L L G G Y Q + S + V T V G P L L K + M I S L F H V + F V G T D G F +

45 Sbjct: 122 A N Y I R Y S S C Q I I L L G G Y Y Q P N S E V T V G P L L K E M I S L F H V N R V F V G T D G F N 172

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 995

- 50 A DNA sequence (GBSx1055) was identified in *S.agalactiae* <SEQ ID 3051> which encodes the amino acid sequence <SEQ ID 3052>. Analysis of this protein sequence reveals the following:

Possible site: 24
>>> Seems to have no N-terminal signal sequence

55 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.1672(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

-1101-

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

5 >GP:AAG04879 GB:AE004578 probable transcriptional regulator
[Pseudomonas aeruginosa]
Identities = 20/70 (28%), Positives = 40/70 (56%)

Query: 6 GFMGRDLMRSEVAQEMANADEVIIITDSSKFNQTALVEQLPLSTVSQVITDKHPNSEIA 65
G M + +E+A+ M A ++ ++ DSSK + AL + PLS +++++ D+ P E+

10 Sbjct: 179 GAMDFSIEBAETIARAMIAQARQLTVIADSSKLGRRALFQVFP LSRINRLVDRKPTGELW 238

Query: 66 NLFQAEIITI 75
Q+A + +

15 Sbjct: 239 EALQQARVEV 248

There is also homology to SEQ ID 3050.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 996

20 A DNA sequence (GBSx1056) was identified in *S.galactiae* <SEQ ID 3053> which encodes the amino acid sequence <SEQ ID 3054>. This protein is predicted to be transcriptional regulator. Analysis of this protein sequence reveals the following:

Possible site: 46
>>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.0904 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

30 A related GBS nucleic acid sequence <SEQ ID 9541> which encodes amino acid sequence <SEQ ID 9542> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

35 >GP:BAB04499 GB:AP001509 transcriptional regulator [Bacillus halodurans]
Identities = 98/309 (31%), Positives = 178/309 (56%), Gaps = 1/309 (0%)

Query: 6 ERQKLLAKVAYLYYMEGKSQSEIANELGIYRTTISRMLAKAREEGLVRIEISDFNPEIFQ 65
E ++L+ KVA LYY EG +Q+++A ++G+ R IS++L KA+E+G+V I I D N +

40 Sbjct: 5 BERRLIVKVASLYYFEGWTQAQVAKKIGVSRPVISKLNKAKEQGIVEIYIKDENIHTVE 64

Query: 66 LESYFKSKYHLKDIEIVSSRKSDTSEIEKDLAHVAAAMIRKKIKENDKVGIAWGRTLSK 125
LE + KYHLK+ +V + I++ + + + K IK D +GI+WG T+S

Sbjct: 65 LEQRLEKKYHLKEAIVVPT-SGLTQDMIKRAIGKATSYVSKNIKGMSIGISWGTTVSS 123

45 Query: 126 VVEAMRPHPVSVQVSFVPLAGGPHSHINARYHVNTLVYEMSRRFQGSCTFINATLVQENANL 185
V+ ++ +PL GG H N L YE++++ C+++ A + E L

Sbjct: 124 FVQEYPYEQHRELKVIPLVGGMGGRKFVELHSNLLAYELAKMNCRCESYLYAPAMVEAKEL 183

Query: 186 AKGILTSKYFEGLMNDWEKLDVAIVGVGKPKSNEQQWLDLLNQDDFQCLDEEAAVGEIT 245
+ ++ S+ +++ + +A+VG+G K + + ++ L ++D L + AVG+++

50 Sbjct: 184 KERLIQSEDIASVLEEGNRNVKMAVVGIGSPFKGSTMVKNYLKEEDIATLKKIGAVGDMS 243

Query: 246 CRFFNHSGDFVNQHLAKRTIGITILEQLQKVPNRIAVAHGNYKAAALLAVLKKGYINHLVT 305
RF++ G P++ L + IGI L++L+++P I V+ G +K ++ A LK GY++ LVT

55 Sbjct: 244 SRFYDALGQFIDHPLNELVIGIDLDLKRIPVIGVSEGAHKVDSVEAALKGGYLDVLVT 303

Query: 306 DFSTALNIL 314

-1102-

D STA +++
Sbjct: 304 DDSTAQSLI 312

- 5 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3055> which encodes the amino acid sequence <SEQ ID 3056>. Analysis of this protein sequence reveals the following:

Possible site: 44
>>> Seems to have no N-terminal signal sequence

- 10 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.2123 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

- 15 Identities = 165/324 (50%), Positives = 238/324 (72%), Gaps = 1/324 (0%)
- Query: 3 MKLERQKLLAKVAYLYMEGKSQSEIANELGIYRTTISRMLAKAREEGLVRIEISDFNPE 62
MK ER++LLAKVAYL+Y++GKSQ+ I+ E+ IYRTT+ RMLAKA+EEG+VRIEI+D++ +
Sbjct: 1 MKEERRRLAKVAYLHYVQGKSQTLSKEMNIYRTTVCRMLAKAKEEGIVRIEILADYDAD 60
- 20 Query: 63 IFQLESYFKSKYHLKDIEIVSSRKSDTSEIEKDLAHVAAAMIRKKIKENDKVGIAWGR 122
+F LE Y + +Y L+ +++V ++ + + ++A AA + R +K+ DK+G++WG T
Sbjct: 61 LFALEEYVRQQYGLEKLDLVPNQVEDTPMDTLTNVAKTAAEVFRHVVDGDKIGLSWGAT 120
- 25 Query: 123 LSKVVEAMRPHPVSVQVSFVPLAGGPPSHINARYHVNTLVYEMSRRFQGSCTFINATLVQEN 182
LS +++ + P + V PLAGGPPSHINA+YHVNTLVY ++R F G+ F+NA ++QE+
Sbjct: 121 LSCMLDELNPKAMKDVFTYPLAGGPPSHINAKYHVNTLVYRLARIFHGNSAFMNAMVIQED 180
- 30 Query: 183 ANLAKGILTSKYFEGLMNDWEKLDVAIVGVGGPKPSNEQ-QWLDLLNQDDFQCLDEEAAV 241
+LAKGIL SKYF ++ +W++LD+A+VG+GG+P S EQ QW DLL D L E AV
Sbjct: 181 KHLAKGILQSKYFNDILTSDWQLDLALVGIGCEPNSLEQSQRDLLTSSDHDQLKYEKAV 240
- 35 Query: 242 GEITCRFFNHSGDPVNOHLAKRTIGITLQLQKVPNRIVAHGNYKAAALLAVLKKGYIN 301
GE+ CRFF+ +G PV L RTIGI+LEQL++VP +AVA G +KA A+LA LK G+IN
Sbjct: 241 GEVCCRFFDQAGQPVYTGLDRTIGISLEQLRRVPKTMVAVATGKHAKAILAALKAGFIN 300
- 40 Query: 302 HLVTFDFSTALNLRDLKDTFVDTI 325
+LVTD T L +L LD+D ++ +
Sbjct: 301 YLVTDKETMLAVLALDEDEDLNNV 324

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 997

- 45 A DNA sequence (GBSx1057) was identified in *S.agalactiae* <SEQ ID 3057> which encodes the amino acid sequence <SEQ ID 3058>. This protein is predicted to be PTS enzyme III cel (celC). Analysis of this protein sequence reveals the following:

Possible site: 55
>>> Seems to have a cleavable N-term signal seq.

- 50 ----- Final Results -----
bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

- 55 A related GBS nucleic acid sequence <SEQ ID 9543> which encodes amino acid sequence <SEQ ID 9544> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

-1103-

>GP:AAA23551 GB:M93570 PTS enzyme III cel [Escherichia coli]
Identities = 42/102 (41%), Positives = 70/102 (68%)

5 Query: 4 EIIVADQIIMGLILNAGDAKQHIYQALKLAKEGNFAESKIEIELADSALLEAHNLQTQFL 63
E+ +++MGLI+N+G A+ Y ALK AK+G+FA +K ++ + AL EAH +QT+ +
Sbjct: 13 EVEELEEVVMGLIINSQARSLAYAALKQAKQGDFAAAKAMMDQSRMALNEAHLVQTKLI 72

Query: 64 AQEAGGTRTDISALFIHSQDHLMTSITEINLIKEIIDLRQEL 105
+AG + +S + +H+QDHLMT+ LI E+I+L ++L
10 Sbjct: 73 EGDAGEGKMKVSLVLVHAQDHLMTSMLARELITELIELHEKL 114

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3059> which encodes the amino acid sequence <SEQ ID 3060>. Analysis of this protein sequence reveals the following:

15 Possible site: 17
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
20 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

25 >GP:AAC74806 GB:AE000268 PEP-dependent phosphotransferase enzyme III
for cellobiose, arbutin, and salicin [Escherichia coli]
Identities = 39/97 (40%), Positives = 66/97 (67%)

Query: 7 DQIIMGLILNAGDAKQHIYQALKCAKEDDYATSEKEMALADDALLEAHNLQTQFLAQEAS 66
+++MGLI+N+G A+ Y ALK AK+ D+A ++ M + AL EAH +QT+ + +A
30 Sbjct: 18 BEVVMGLIINSQARSLAYAALKQAKQGDFAAAKAMMDQSRMALNEAHLVQTKLIEGDAG 77

Query: 67 GNKSEITALFVHSQDHLMTTITEINLIKEIIDLRKEL 103
K +++ + VH+QDHLMT++ LI E+I+L ++L
Sbjct: 78 EGKMKVSLVLVHAQDHLMTSMLARELITELIELHEKL 114

35 An alignment of the GAS and GBS proteins is shown below.

Identities = 81/103 (78%), Positives = 94/103 (90%)

Query: 3 MEIIVADQIIMGLILNAGDAKQHIYQALKLAKEGNFAESKIEIELADSALLEAHNLQTQF 62
M++IV DQIIMGLILNAGDAKQHIYQALK AKE ++A S+ E+ LAD ALLEAHNLQTQF
40 Sbjct: 1 MQVIVPDQIIMGLILNAGDAKQHIYQALKCAKEDDYATSEKEMALADDALLEAHNLQTQF 60

Query: 63 LAQEAGGTRTDISALFIHSQDHLMTSITEINLIKEIIDLRQEL 105
LAQEA G +++I+ALF+HSQDHLMT+ITEINLIKEIIDLR+EL
45 Sbjct: 61 LAQEASGNKSEITALFVHSQDHLMTTITEINLIKEIIDLRKEL 103

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 998

50 A DNA sequence (GBSx1058) was identified in *S.agalactiae* <SEQ ID 3061> which encodes the amino acid sequence <SEQ ID 3062>. This protein is predicted to be PTS system, cellobiose-specific IIB component (celA). Analysis of this protein sequence reveals the following:

Possible site: 24
>>> Seems to have a cleavable N-term signal seq.

55 ----- Final Results -----
bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

-1104-

The protein has homology with the following sequences in the GENPEPT database.

```

5  >GP:AAF94440 GB:AE004207 PTS system, cellobiose-specific IIB
    component [Vibrio cholerae]
    Identities = 46/100 (46%), Positives = 62/100 (62%)

Query: 1  MIKIGLFCAAGFSTGMLVNNMKIAADKEGIEAHIEAYSQGGKIADYAKDLVDVALLGFPQVSY 60
          M KI L C+AG ST MLV M+ AA+ +GIE I+A S      + ++ DV LLGPQV +
10  Sbjct: 1  MKKILLCCSAGMSTSMILVKKMQAAESKGIECKIDALSVNAFEEAIQEYDVCLLGPQVRF 60

Query: 61  TLDKSKSICDEYGVPIAVIPMADYGMLDGKVKLKLALSLL 100
          L++ +  DEYG IA I  YGM+ G +VL+ AL L+
15  Sbjct: 61  QLEELRKTADYEGKNIAAISPOAYGMMKGDEVLLQQAALDLI 100

```

15 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3063> which encodes the amino acid sequence <SEQ ID 3064>. Analysis of this protein sequence reveals the following:

```

    Possible site: 31
    >>> Seems to have a cleavable N-term signal seq.

20  ----- Final Results -----
          bacterial outside --- Certainty=0.3000(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

25 The protein has homology with the following sequences in the databases:

```

    >GP:AAF94440 GB:AE004207 PTS system, cellobiose-specific IIB
    component [Vibrio cholerae]
    Identities = 43/100 (43%), Positives = 58/100 (58%)

30  Query: 8  MIKIGLFCAAGFSTGMLVNNMKVAAEKKGIDCQIEAYAQQGLADYAPLLDVALLGFPQVAY 67
          M KI L C+AG ST MLV M+ AA+ KGI+C+I+A +      +  DV LLGPQV +
    Sbjct: 1  MKKILLCCSAGMSTSMILVKKMQAAESKGIECKIDALSVNAFEEAIQEYDVCLLGPQVRF 60

Query: 68  TLDKSEAICKDNDIPIAVIPMADYGMLDGKVKLKLALSILV 107
          L++ +  IA I  YGM+ G++VL AL L+
35  Sbjct: 61  QLEELRKTADYEGKNIAAISPOAYGMMKGDEVLLQQAALDLI 100

```

An alignment of the GAS and GBS proteins is shown below.

```

    Identities = 79/101 (78%), Positives = 92/101 (90%)

40  Query: 1  MIKIGLFCAAGFSTGMLVNNMKIAADKEGIEAHIEAYSQGGKIADYAKDLVDVALLGFPQVSY 60
          MIKIGLFCAAGFSTGMLVNNMK+AA+K+GI+ IEAY+QGG+ADYA LDVALLGFPQV+Y
    Sbjct: 8  MIKIGLFCAAGFSTGMLVNNMKVAAEKKGIDCQIEAYAQQGLADYAPLLDVALLGFPQVAY 67

45  Query: 61  TLDKSKSICDEYGVPIAVIPMADYGMLDGKVKLKLALSILE 101
          TLDKS++IC +  +PIAVIPMADYGMLDG KVL LALSIL++
    Sbjct: 68  TLDKSEAICKDNDIPIAVIPMADYGMLDGKVKLKLALSILVK 108

```

50 SEQ ID 3062 (GBS180) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 39 (lane 4; MW 12.6kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 41 (lane 2; MW 37.6kDa).

The GBS180-GST fusion product was purified (Figure 204, lane 8) and used to immunise mice. The resulting antiserum was used for FACS (Figure 298), which confirmed that the protein is immunoaccessible on GBS bacteria.

55 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1105-

Example 999

A DNA sequence (GBSx1059) was identified in *S.galactiae* <SEQ ID 3065> which encodes the amino acid sequence <SEQ ID 3066>. This protein is predicted to be pts system, cellobiose-specific iic component (celB). Analysis of this protein sequence reveals the following:

```

5      Possible site: 40
      >>> Seems to have no N-terminal signal sequence
      INTEGRAL      Likelihood = -11.68      Transmembrane      346 - 362 ( 334 - 374)
      INTEGRAL      Likelihood = -9.77       Transmembrane      182 - 198 ( 178 - 205)
10     INTEGRAL      Likelihood = -8.65       Transmembrane      29 - 45 ( 27 - 50)
      INTEGRAL      Likelihood = -6.53       Transmembrane      140 - 156 ( 134 - 161)
      INTEGRAL      Likelihood = -4.78       Transmembrane      292 - 308 ( 289 - 312)
      INTEGRAL      Likelihood = -4.41       Transmembrane      397 - 413 ( 395 - 416)
      INTEGRAL      Likelihood = -2.97       Transmembrane      77 - 93 ( 72 - 93)
15     INTEGRAL      Likelihood = -2.97       Transmembrane      228 - 244 ( 222 - 246)

      ----- Final Results -----
      bacterial membrane --- Certainty=0.5670(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
20     bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

      >GP:AAA17390 GB:U07818 cellobiose phosphotransferase enzyme II''
      [Bacillus stearothermophilus]
      Identities = 160/415 (38%), Positives = 251/415 (59%), Gaps = 13/415 (3%)
25     Query: 15  KFMNMRGIIALKDGM LAILPLTVVGSFLILGQLPFPKGLNQAIANVFGPEWTEPFM QVYS 74
      K   R + A++DG++ +PL ++GSLFLI+G LP G N+ +A FG W + +
      Sbjct: 18  KIAEQRHLQAIRDGIILSMPLLIIGSLFLIVGFLPIPGYNEWMAKWFGEHWLDKLLYPVG 77

30     Query: 75  GTFPAIMGLISCF AIAIYAYAKNSSVEPLPAGVLSLSSFFILMKSSYIPVKGEA-----IA 128
      TF IM L+ F +AY A+ V+ L AG +SL++F +L +P E ++
      Sbjct: 78  ATFDIMALVVSFGVAYRLAEKYKVDALSAGAISLA AF-LLATPYQVPFTPEGAKETIMVS 136

35     Query: 129 DAISKVWFQGGQGIIGAI IIGLVVGAITYTWF IQHHIVIKMPEQVPQAIKQFEAMIPAFVI 188
      I   W G +G+ A+I+ +V IY IQ +IVIK+P+ VP A+A+ F A+IP +
      Sbjct: 137 GGI PVQWVGSKGLFVAMILAIVSTETIYRKIIQKNIVIKLPDGVPPAVARSFVALIPGA AV 196

40     Query: 189 FLLSMIVYLI AKVTTGGTFIEMIYDIIQVPLQGLTGS LYG AIGIAFFISFLWVFGVHGQS 248
      ++ + LI ++T +F ++ ++ PL L GS++GAI + LW G+HG +
      Sbjct: 197 LVVVVVARLILEMTPPFESFHNIVSVLLNKPLSVLGGSVFGAIVAVLLVQLLWSTGLHGAA 256

45     Query: 249 VVNGIVTALLLSNLDANKSLLAAN-RITLDNGAHIVTQQFLDSFLILSGSGITFGLVIAM 307
      +V G++ + LS +D N+ + N L N ++TQQF D ++ + GSG T L + M
      Sbjct: 257 IVGVMGP IWL SLM DENRMVFQQNPN AELN---VITQQFFDLWIYIGSGGATLALALTM 313

50     Query: 308 LFAAKSKQYKALGKVA AFPAIFNVNEPIVFGFPITVMNPVMFLPFILVPVLAALIVYGAIA 367
      +F A+S+Q K+LG++A P IFN+NEPI FG PIVMNP++ +PFILVPV+ ++ Y A+A
      Sbjct: 314 MFRARSRLKSLGRLAIAPGIFNINEPITFGMPITVMNPLLIIPFILVPVVLVVVSYAAMA 373

55     Query: 368 VGFMQPFSGVTLPWSTPAIISGFMVGGWQ--GALVQIVILAISTAVYFPFFKIQD 420
      G + SGV +PW+TP +ISG++ G + G+++QIV I+ A+Y+PFF I D
      Sbjct: 374 TGLVAKPSGVAVPWTTPITVISGYLATGGKISGSILQIVNFFIAFATYPPFFSIWD 428

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2215> which encodes the amino acid sequence <SEQ ID 2216>. Analysis of this protein sequence reveals the following:

```

      Possible site: 40
      >>> Seems to have no N-terminal signal sequence
      INTEGRAL      Likelihood = -8.92       Transmembrane      347 - 363 ( 335 - 373)
      INTEGRAL      Likelihood = -7.59       Transmembrane      29 - 45 ( 27 - 50)
60     INTEGRAL      Likelihood = -7.38       Transmembrane      182 - 198 ( 179 - 204)
      INTEGRAL      Likelihood = -5.68       Transmembrane      398 - 414 ( 395 - 420)
      INTEGRAL      Likelihood = -4.99       Transmembrane      293 - 309 ( 291 - 314)

```

-1106-

```

INTEGRAL    Likelihood = -3.61    Transmembrane  140 - 156 ( 134 - 160)
INTEGRAL    Likelihood = -2.60    Transmembrane  229 - 245 ( 229 - 246)
INTEGRAL    Likelihood = -0.75    Transmembrane   72 -  88 (  72 -  88)

```

5

----- Final Results -----

```

bacterial membrane --- Certainty=0.4567(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

10 An alignment of the GAS and GBS proteins is shown below.

Identities = 366/428 (85%), Positives = 402/428 (93%), Gaps = 1/428 (0%)

```

Query: 1  MSKFDSQKIITPIMKFNMRGIILKDGMLAILPLTVVGSFLFLILGQLPFFKGLNQAIANV 60
M+K + Q II PIM FVNMRIILKDGMLAILPLTVVGSFLFLI GQ+PF+G+N AIA+V
15 Sbjet: 1  MAKNMNQNIIPIMTFVNMRIILKDGMLAILPLTVVGSFLFLIAGQIPFQGVNDIAISV 60

Query: 61  FGPEWTEPFMQVYSGTFAIMGLISCFATAYAYAKNSSVEPLPAGVLSLSSFFILMKSSYI 120
FG +WTEPFMQVY GTFAIMGLISCFAT Y+YAKNS VEPLP+GVLSLS+FFIL++SSY+
20 Sbjet: 61  FGADWTEPFMQVYHGTFAIMGLISCFATIGYSYAKNSGVEPLPSGVLSLSAFFILLRSSYV 120

Query: 121  PVKGEAIDADAISKVWFGGQGIIGAIIGLVVGAITYTFIQHHIVIKMPEQVPQAIKQFE 180
P +GEAI DAISKVWFGGQGIIGAI+IGL VGA+YT FI+ HIVIKMP+QVPQAIKQFE
25 Sbjet: 121  PAEGEAIGDAISKVWFGGQGIIGAIIVIGLTVGAVYTTFIRRHIVIKMPDQVPQAIKQFE 180

Query: 181  AMIPAFVIFLLSMIVYLIK-VTTGGTFIEMIYDIIQVPLQGLTGSLYGAIGIAFFISFL 239
AMIPAFVIF LSM+VY+IAK VT GGTFIEMIYD+IQVPLQGLTGSLYGA+GIAFFISFL
30 Sbjet: 181  AMIPAFVIFLLSMLVYIIAKSVTGGTFIEMIYDVIQVPLQGLTGSLYGALGIAFFISFL 240

Query: 240  WWFGVHGQSVVNGIVTALLLSNLDANKSLLAANRLTLDNGAHIVTQQFLDSFLILSGSGI 299
WWFGVHGQSVVNGIVTALLLSNLDANK+L+AA L+LD GAHIVTQQFLDSFLILSGSGI
35 Sbjet: 241  WWFGVHGQSVVNGIVTALLLSNLDANKALMAAGELSLDKGAHIVTQQFLDSFLILSGSGI 300

Query: 300  TFGLVIAMLFAAKSKQYKALGKVAAPFAIFNVNEPIVFGFPIVMNPVMFLPFILVPLVLA 359
TFGLV+AM+FAAKSKQYKALGKVAAPFA+FNVNPE+VFGFPIVMNPVMFLPFILVPLVLA
40 Sbjet: 301  TFGLVVAAMIFAAKSKQYKALGKVAAPFALFNVNPEVVFPGFPIVMNPVMFLPFILVPLVLA 360

Query: 360  LIVYGAIAGVFMQPFSGVTLPWSTPAIISGFMVGGWQGAIVQIVILAISTAVYFPFFKIQ 419
L VYGAIA+GFMQPF+GVTLPWSTPAIISGFMVGGWQGA+VQI+IL +ST VYFPFFKIQ
45 Sbjet: 361  LTVYGAIAGVFMQPFAGVTLPWSTPAIISGFMVGGWQGAIVQIILILIMSTLVYFPFFKIQ 420

Query: 420  DNITYKNE 427
DN+ Y+NE
50 Sbjet: 421  DNMAVQNE 428

```

45 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1000

A DNA sequence (GBSx1060) was identified in *S. agalactiae* <SEQ ID 3067> which encodes the amino acid sequence <SEQ ID 3068>. This protein is predicted to be formate acetyltransferase 2 (pflB). Analysis of this protein sequence reveals the following:

```

Possible site: 54
>>> Seems to have no N-terminal signal sequence

```

55

----- Final Results -----

```

bacterial cytoplasm --- Certainty=0.5049(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

60

```

>GP:AAC73910 GB:AE000184 putative formate acetyltransferase

```

-1107-

[Escherichia coli K12]

Identities = 414/805 (51%), Positives = 555/805 (68%), Gaps = 14/805 (1%)

5 Query: 25 LTERMYSYRDKVLD-KKPFIDAERAILVTEAYQKHQEKPNVLKRAYMLQNILEKMTIYID 83
L++R+ ++++ ++ KP + ERA TE YQ+H +KP ++RA L + L TI+I
Sbjct: 9 LSDRIKAHKNALVHIVKPPVCTERAQHYTEMYQQHLDKPIPVRRALALAHHLANRTIWIK 68

10 Query: 84 DETMIVGNQASSDKDAPIFPEYTFLEFVVNELDLFEKRDGDFYITEETKEQIRNIAPFWE 143
+ +I+GNQAS + APIFPEYT+ ++ E+D R G F ++EE K + + P+W
Sbjct: 69 HDELIIGNQASEVRAAPIFPEYTVSWIEKEIDDLADRPAGFAVSEENKRVLHEVCPWWR 128

15 Query: 144 NNNLRARAGVMLPEEVQVYMETGFFGMEGKMNSGDAHLAVNYQKLLLEEGLIGFEKKARKA 203
++ R M +E + + TG EG M SGDAHLAVN+ LLE+GL G ++ +
Sbjct: 129 GQTVDRCYGMFTDEQKGLLATGIIKAEGNMTSGDAHLAVNFPLLLEKGLDCLREEVAER 188

20 Query: 204 KADLDLTKPESIDKYHFYSILITIEAVKYAERFAILAKKQAKTANAK-RRQELLDIAS 262
++ ++LT E + F +I I + AV + ERFA LA++ A T + RR ELL +A
Sbjct: 189 RSRINLTVLEDLHGEQFLKAIDIVLVAVSEHIERFAALAREMAATETRESRDELLAMAE 248

25 Query: 263 ICERVPPYPAETFAEAVQSVWFIFQICILQIESNGHSLSYGRFDQYMPYVKSDELAGRETE 322
C+ + + P +TF +A+Q +FIQ ILQIESNGHS+S+GR DQY+YPY + D+E + +
Sbjct: 249 NCDLIAHQPPQTFWQALQCYFIQLILQIESNGHSVSFGRMDQYLYPYRRDVELNQTLTD 308

30 Query: 323 -DSIVERLTNLWIKTITINKVRSQAHTFSSAGSPLYQNVITIGGQTR---HKEDAVNPLSF 378
+ +E L + W+K + +NK+RS +H+ +SAGSPLYQNVITIGGQ DAVNPLS+
Sbjct: 309 REHAIEMLHSCWLKILEVNKIRSGSHSKASAGSPLYQNVITIGGQNLVDGQPMDAVNPLSY 368

35 Query: 379 LVLKSVAQTHLPQPNLTVRYHANLDSFMNEAIEVMKLGFGMPAFNNDIIPFIKKG 438
+L+S + QPNL+VRYHA + F++ ++V++ GFGMPAFNNDIIP FIK G+
Sbjct: 369 AILESCGRLRSTQPNLSVRYHAGMSNDFLDACVQVIRCGFGMPAFNNDIIVPEFIKGI 428

40 Query: 439 SEEDAYDYSAIGCVETAVPGKWGYRCTGMSYINFPKVLITMNDGIDPASGKRFPAP---- 494
+DAYDY+AIGC+ETAV GKWGYRCTGMS+INF +V+L + G D SGK F P
Sbjct: 429 EPQDAYDYAAIGCIETAVGGKWGYRCTGMSFINFARVMLAALEGCHDATSGKVFLPQEKA 488

45 Query: 495 -SYGHFTQMTSYKELKEAWDKTLRYLTRMSVIVENAIDISLEREVPDILCSALTDDCIGR 553
S G+F ++ E+ +AWD +RY TR S+ +E +D LE V DILCSAL DDCI R
Sbjct: 489 LSAGNFN---NFDEVMDAWDTQIRYTRKSIEIEFYVDTMLEENVHDILCSALVDDCIER 545

50 Query: 554 GKHLKEGGAVDYDISGLQVGIANLSDSLAALKKLVFEKRLITLLEVWQALQSDYAGPRGE 613
K +K+GGA YD++SGLQVGIANL +SLAA+KKLVFE+ + ++ AL D+ G E
Sbjct: 546 AKSIKQGGAKYDWWVSGLQVGIANLGNLSAAVKKLVFEQGAIGQQQLAALADDFDGLTHE 605

55 Query: 614 EIRQMLINEAPKYGNDDYADSLVRECYDVYVEEIAKYPNTRYGRGPPIGGIRYSGTSSIS 673
++RQ LIN APKYGNDD D+L+ Y Y++E+ +Y N RYGRGP+GG Y+GTSSIS
Sbjct: 606 QLRQRLINGAPKYGNDDTVDTLARAYQTYIDELKQYHNPRYGRGPVGGNYAGTSSIS 665

60 Query: 674 ANVGQGRGTLATPDGRHAGTPLAEGCSPSHNMDKKGPTSVLKS VSKLPTDEIVGGVLLNQ 733
ANV G T+ATPDGR A TPLAEG SP+ D GPT+V+ SV KLPT I+GGVLLNQ
Sbjct: 666 ANVPFGAQTMAPDGRKAHTPLAEGASPASGT'DHLGPTAVIGSVGKLPTAAILGGVLLNQ 725

65 Query: 734 KVNPTLAKEEDKQKLIALLRTFFNRLHGYHIQYNVVSRETLIDAQKHPEKHRDLIVRVA 793
K+NP TL E DKQKL+ LLRTFF G+HIQYN+VSRETL+DA+KHP+++RDL+VRVA
Sbjct: 726 KLNPATLENESDKQKLMILLRTFFEVHKGWHIQYNIVSRETLIDAKKHPDQYRDLVVRVA 785

Query: 794 GYSAFFNVLSKATQDDIIARTEHAL 818
GYSAFF LS QDDIIARTEH L
Sbjct: 786 GYSAFFTALSPDAQDDIIARTEHML 810

60 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3069> which encodes the amino acid sequence <SEQ ID 3070>. Analysis of this protein sequence reveals the following:

Possible site: 19

>>> Seems to have no N-terminal signal sequence

65 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.4763 (Affirmative) < succ>

-1108-

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

5 Identities = 694/803 (86%), Positives = 747/803 (92%)

Query: 16 QNSQKHFGYLTERMSYRDKVLDDKKPFIDAERAILVTEAYQKHQKPNVLKRAYMLQNIL 75
 + +FG+LT+RM YR+ VLDKKP+IDAERAIL TEAYQKHQ KP LKRAYMLQ IL
 10 Sbjct: 3 ETKSPYFGHLTDRMTHYREAVLDKKPYIDAERAILATEAYQKHQKPNLKRAYMLQITIL 62

Query: 76 EKMTIYIDDETMIVGNQASSDKDAPIFPEYTLFVFNELDLFEKRDGDVIFYITEETKEQI 135
 E MTIYI+DE++I GNQASS+KDAPIFPEYTLFV+NELDLFEKRDGDVIFYITEETK+Q+
 15 Sbjct: 63 ENMTIYIEDESLIAGNQASSNKAIFPEYTLFVFNELDLFEKRDGDVIFYITEETKQQL 122

Query: 136 RNIAPFWENNLRARAGVMLPEEVQVYMETGFFGMEGKMNSGDAHLAVNYQKLEELGLIG 195
 R+IAPFWENNLRAR GV+LPEEVQVYMETGFFGMEGKMNSGDAHLAVNYQKLE GL G
 20 Sbjct: 123 RDIAPFWENNLRARCGVLLPEEVQVYMETGFFGMEGKMNSGDAHLAVNYQKLEHGLKG 182

Query: 196 FEKKARKAKADLDLTKPESIDKYHFYDSILITIEAVKTYAERFAILAKKQAKTANAKRRQ 255
 FE++AR AKA LDLT PE+IDKYHFYDS+ I I+AVKTYA+R+A LA++ AKTA +R+
 25 Sbjct: 183 FEERARAACAALDLTIPENIDKYHFYDSVFIVIDAVKTYAKRYAKLARELAKTAKPERQA 242

Query: 256 ELLDIASICERVYPYPAETFAEAVQSVWFIQCILQIESNGHSLSYGRFDQYMPYVKSDDL 315
 ELLDIA IC++VPY PA+TFAEAVQSVWFIQCILQIESNGHSLSYGRFDQYMPYVK+DL
 30 Sbjct: 243 ELLDIARICDKVPYEPAKTFAEAVQSVWFIQCILQIESNGHSLSYGRFDQYMPYVKADL 302

Query: 316 EAGRETEDSIVERLTNLWIKTITINKVRSQAHTFSSAGSPLYQNVITIGGQTRHKEDAVNP 375
 EAGRETED+IVERLTNLWIKT+TINKVRSQAHTFSSAGSPLYQNVITIGGQTR K+DAVNP
 35 Sbjct: 303 EAGRETEDTIVERLTNLWIKTLTINKVRSQAHTFSSAGSPLYQNVITIGGQTRDKKDAVNP 362

Query: 376 LSFLVLKSAQTHLPQPNLTVRYHANLDSFMNEAIEVMKLGFGMPAFNNDEIIPSFIK 435
 LS+LVL+SVAQT LPQPNLTVRYH LD +FMNE IEVMKLGFGMPA NNDEIIPSFIK
 40 Sbjct: 363 LSYLVLRSAQTKLPQPNLTVRYHKGDLNTFMNECIEVMKLGFGMPAMNNDEIIPSFIK 422

Query: 436 KGVSEEDAYDYSAIGCVETAVPGKWGYRCTGMSYINFPKVLITMNDGIDPASGKRFAPS 495
 KGVSEEDAYDYSAIGCVETAVPGKWGYRCTGMSYINFPK+LLITMNDGIDPASGKRFA
 45 Sbjct: 423 KGVSEEDAYDYSAIGCVETAVPGKWGYRCTGMSYINFPKILLITMNDGIDPASGKRFAK 482

Query: 496 YGHFTQMTSYKELKEAWDKTLRYLTRMSVIVENAIDISLEREVPDILCSALTDDCIGRGK 555
 +GHF MTSY+ELK AWD TLR +TRMSVIVENAID+ LEREVPDILCSALTDDCIGRGK
 50 Sbjct: 483 HGHFKDMTSYEELKAAWDA'LTREITRMSVIVENAIDLGLEREVPDILCSALTDDCIGRGK 542

Query: 556 HLKEGGAVIDYISGLQVGIANLSDSLAALKKLVFEEKRLTLEWVQALQSDYAGPRGEEI 615
 LKEGGAVIDYISGLQVGIANLSDSLAALKKLVFEE RLT E+W+AL+SD+AG RGE+I
 55 Sbjct: 543 TLKEGGAVIDYISGLQVGIANLSDSLAALKKLVFEEGRLTPEELWKALESDFAGERGEDI 602

Query: 616 RQMLINEAPKYGNDDYADSLVRECYDVVEETAKYPNTRYGRGPIGGIRYSGTSSISAN 675
 RQMLIN+APKYGNDDYADSLV E YD Y++EIAKYPNTRYGRGPIGGIRYSGTSSISAN
 60 Sbjct: 603 RQMLINDAPKYGNDDYADSLVVEAYDITYDEETAKYPNTRYGRGPIGGIRYSGTSSISAN 662

Query: 676 VGQGRGTLATPDGRHAGTPLAEGCSPSHNMDKKGPTSVLKS SVSKLPTDEIVGGVLLNQKV 735
 VGQG+GTLATPDGRHAGTPLAEGCSP H+MDKKGPTSVLKS V+KLPTDEIVGGVLLNQKV
 65 Sbjct: 663 VGQKGTLATPDGRHAGTPLAEGCSPSHNMDKKGPTSVLKS VAKLPTDEIVGGVLLNQKV 722

Query: 736 NPQTLAKEEDKQKLIALLRTFFNRLHGYHIQYNNVVSRETLLDAQKHPEKHRDLIVRVAGY 795
 NPQTLAKEEDK KL+ALLRTFFNRLHGYHIQYNNVVSRETLLDAQKHPEKHRDLIVRVAGY
 70 Sbjct: 723 NPQTLAKEEDKLLKLMALLRTFFNRLHGYHIQYNNVVSRETLLDAQKHPEKHRDLIVRVAGY 782

Query: 796 SAFFNVLSKATQDDIARTHEAL 818
 SAFFNVLSKATQDDII RTEH L
 75 Sbjct: 783 SAFFNVLSKATQDDIERTHTL 805

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1109-

Example 1001

A DNA sequence (GBSx1061) was identified in *S.agalactiae* <SEQ ID 3071> which encodes the amino acid sequence <SEQ ID 3072>. Analysis of this protein sequence reveals the following:

```

Possible site: 32
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.1024(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:CAA05516 GB:AJ002527 OrfX [Clostridium beijerinckii]
Identities = 90/214 (42%), Positives = 131/214 (61%), Gaps = 1/214 (0%)

Query: 1  MEFLDLTLNLEAIKKWHHILPLAGVTSNPTIAKKEGDIHFFQRI RDVREIIGREASLHVQ 60
      M+ ++D +N+E IK   I + GVTSNP+I K G +   I+ +RE IG + LHVQ
Sbjct: 1  MKLIIDDVNIEKIKDVFSIFQIDGVTSNPSILHKYKQPYEILIK-IREFIGENSELHVQ 59

Query: 61  VVAKDYQGILDDAAKIRQETDDDIYIKVPVTPDGLAAIKTLKAEGYNITATAIYTSMQGL 120
      V+++ +G+L +A KI +E   + Y+K+PVT DGL AIK L+ E N+TATAIYT MQ
Sbjct: 60  VISESSEGMLKEAHKIIKELGKNYVVKIPVTRDGLKAIKILRKEINV TATAIYTMQAY 119

Query: 121 LAISAGADYLAPYFNRMENLDIDATQVIKELAQAIERTGSSSKILAA SFKNASQVTKALS 180
      LA AGA Y APY NR++NL + QV K++   E+   +++LAASFKN+ QV +
Sbjct: 120 LAGKAGAQAAPYVNRIDNLGANGVQVAKDIHDI FEKNNFKEVLAASFKN SQVLELCK 179

Query: 181 QGAQSITAGPDIFESVFAMP SIAKAVNDFADDWK 214
      G + T PD+ E +   + AV +F D++
Sbjct: 180 YGIGAATISPDVIEGLIKNDCVDVA VENFKKDFE 213

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3073> which encodes the amino acid sequence <SEQ ID 3074>. Analysis of this protein sequence reveals the following:

```

Possible site: 32
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.1090(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

```

Identities = 158/222 (71%), Positives = 194/222 (87%)

Query: 1  MEFLDLTLNLEAIKKWHHILPLAGVTSNPTIAKKEGDIHFFQRI RDVREIIGREASLHVQ 60
      ME++LDTL+LEAIKKWHHILPLAGVTSNP+IAKKEG+I FF+RIR+VR IIG +AS+HVQ
Sbjct: 1  MEYMLDTLDLEAIKKWHHILPLAGVTSNPSIAKKEGEIDFFERIREVRAIIGDKASIHVQ 60

Query: 61  VVAKDYQGILDDAAKIRQETDDDIYIKVPVTPDGLAAIKTLKAEGYNITATAIYTSMQGL 120
      V+A+DY+GIL DAA+IR++ D +Y+KVPVT +GLAAIKTLKAEGY+ITATAIYT+ QGL
Sbjct: 61  VIAQDYEGILKDAAEIRQC GDSVYKVPVITTEGLAAIKTLKAEGYHITATAIYTTFQGL 120

Query: 121 LAISAGADYLAPYFNRMENLDIDATQVIKELAQAIERTGSSSKILAA SFKNASQVTKALS 180
      LAI AGADYLAPY+NRMENL+ID VI++LA+AI R ++SKILAA SFKN +QV K+ +
Sbjct: 121 LAIEAGADYLAPYVNRMENLNDPEAVIEQLAEATNRENANSKILAA SFKNVAQVNKSFA 180

Query: 181 QGAQSITAGPDIFESVFAMP SIAKAVNDFADDWKASQHSEHI 222
      GAQ+ITAGPD+FE+ FAMP SIAKAV+DF DW+A H + I
Sbjct: 181 LGAQAITAGPDVFEAGFAMP SIQKAVDDFGKDWEAIIHHRKSI 222

```

-1110-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1002

A DNA sequence (GBSx1062) was identified in *S.agalactiae* <SEQ ID 3075> which encodes the amino acid sequence <SEQ ID 3076>. Analysis of this protein sequence reveals the following:

Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

10 bacterial cytoplasm --- Certainty=0.3086(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9545> which encodes amino acid sequence <SEQ ID 9546> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAA22477 GB:M65289 glycerol dehydrogenase [Bacillus
 stearothermophilus]
Identities = 199/362 (54%), Positives = 271/362 (73%), Gaps = 2/362 (0%)
20 Query: 4 KVFASPSRYIQGDALFQSIIEHKSIGQTPILCDDVVYNIVGERFLSYLQD-DLLPHRV 62
 +VF SP++Y+QGK+ + + +++ +G +++ D++V+ I G ++ L+ ++ V
 Sbjct: 5 RVFISPAKYVQGNVITKIANYLEGIGNKTFVVIADIEIVWKIAGHTIVNELKKGNIAAEEV 64
25 Query: 63 SFNGEASDNEINRVVAVAKEKNSDLIIIGLGGGKTIDSAKAIADKVNLPVVIAPTAVASTDA 122
 F+GEAS NE+ R+ +A++ + ++IG+GGGKT+D+AKA+AD+++ +VI PT ASTDA
 Sbjct: 65 VFSGEASRNEVERIANIARKAEAAIIVIGVGGGKTLDTAKAVALDELDAYIVIVPTAASTDA 124
30 Query: 123 PTSALSVIYTDEGAFEKYIFYSKNPDLVLVDVTQVIAQAPKRLASGIADGLATWVEARAV 182
 PTSALSVIY+D+G FE Y FY KNPDLVLVD++IA AP RLLASGIAD LATWVEAR+V
 Sbjct: 125 PTSALSVIYSDDGVFESYRFYKKNPDLVLVDTKIIANAPRLLASGIADALATWVEARSV 184
35 Query: 183 LQKNGIAMAGGRQTLAGVAIAQACERTLFDNSLQALACDAKVVTKALENVIEANTLLSG 242
 ++ G MAGG T+A AIA+ CE+TLF A + AKVVT ALE V+EANTLLSG
 Sbjct: 185 IKSGGKTMAGGIPTIAAEIAAEKCEQTLFKYGLAYESVKAKVVTPEALEAVVEANTLLSG 244
40 Query: 243 LGFESAGLAAAHAIHNGFTALSGDIHHLTHGEKVAYGTLTQLFLENRPKEEIDRYINLYQ 302
 LGFES GLAAAHAIHNGFTAL G+IHHLTHGEKVA+GTL QL LE ++EI+RYI LY
 Sbjct: 245 LGFESGGLAAAHAIHNGFTALEGEIHHLTHGEKVAFGTLVQLALEEHSQQEIERIYIELYL 304
45 Query: 303 AIGMPTTLAELHLGDATYEELLKVGQQTIEGETIHEMPFKISAEDVAAALLTVDRYVSN 362
 ++ +P TL ++ L DA+ E++LKV + AT EGETIH F ++A+DVA A+ D+Y
 Sbjct: 305 SLDLPTLEDIKLDASREDILKVAKAATAEGETIHN-AFNVTTADDVADAIFAADQYAKA 363
55 Query: 363 HQ 364
 ++
 Sbjct: 364 YK 365

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3077> which encodes the amino acid sequence <SEQ ID 3078>. Analysis of this protein sequence reveals the following:

Possible site: 35
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -4.62 Transmembrane 101 - 117 (98 - 119)

----- Final Results -----

 bacterial membrane --- Certainty=0.2848(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

-1111-

The protein has homology with the following sequences in the databases:

>GP:AAA22477 GB:M65289 glycerol dehydrogenase [Bacillus
stearothermophilus]
Identities = 202/357 (56%), Positives = 261/357 (72%), Gaps = 1/357 (0%)

Query: 2 KVFASPSRYIQGKNALFTNVKTLKQLGDSPIILLCDDVVYGVGERFESYLIDNGMTPVHV 61
+VF SP++Y+QGKN + L+ +G+ +++ D++V+ I G + L + V
Sbjct: 5 RVFISPAKYVQGNVITKIANYLEGIGNKTVVIADIEIVWKIAGHTTIVNELKKGNIAAEEV 64

Query: 62 AFNGEASDNEISRVAIAKENGNDVIIGLGGGKTIDSAKAIADLLAVPVIIAPTIASTDA 121
F+GEAS NE+ R+ IA++ ++IG+GGGKT+D+AKA+AD L ++I PT ASTDA
Sbjct: 65 VFSGEASRNEVERIANIARKAEAAIIVIGVGGGKTLDTAKAVALDAYIVIVPTAASTDA 124

Query: 122 PTSALSVIYTDEGAPEKYIFYSKNPDLVLVDVTQVICQAPKRLLASGIADGLATWVEARAV 181
PTSALSVIY+D+G FE Y FY KNPDLVLVD++I AP RLLASGIAD LATWVEAR+V
Sbjct: 125 PTSALSVIYSDDGVFESYRFYKKNPDLVLVDTKIIANAPRLLASGIADALATWVEARSV 184

Query: 182 MQKNGDTMAGGNQTLAGVAIAKACEQTLFADGLKAMASCDRQVVTPALENVIEANTLLSG 241
++ G TMAGG T+A AIA+ CEQTLF G A S +VVTPALE V+EANTLLSG
Sbjct: 185 IKSGGKTMAGGIPTIAAEIAIEKCEQTLFKYKGLAYESVKAKVVTPALEAVVEANTLLSG 244

Query: 242 LGFESAGLAAAHAIHNGFTALTGAIIHHLTHGEKVAYGTLTQLFLENRSREEIDRYIDFYQ 301
LGFES GLAAAHAIHNGFTAL G IHHHLTHGEKVA+GTL QL LE S++EI+RYI+ Y
Sbjct: 245 LGFESGGLAAAHAIHNGFTALEGEIHHHLTHGEKVAFGTLVQLALEEHSQQBIERYIELYL 304

Query: 302 AIGMPTTLKEMHLDATQEDFLKIGRQATMAGETIHQMPFVISPEDVAAALVAVDAY 358
++ +P TL+++ L A++ED LK+ + AT GETIH F ++ +DVA A+ A D Y
Sbjct: 305 SLDLPTLEDIKLDASREDILKVAKAATAEGETIHN-AFNVTADDVADAIFAADQY 360

An alignment of the GAS and GBS proteins is shown below.

Identities = 287/361 (79%), Positives = 325/361 (89%), Gaps = 1/361 (0%)

Query: 3 MKVFASPSRYIQGKDALFQSIIEHIKSLGQTPILLCDDVVYNIVGERFSLYLQDD-LLPHR 61
MKVFASPSRYIQGK+ALF +++ +K LG +P++LCDDVVY IVGERF SYL D+ + P
Sbjct: 1 MKVFASPSRYIQGKNALFTNVKTLKQLGDSPIILLCDDVVYGVGERFESYLIDNGMTPVH 60

Query: 62 VSFNGEASDNEINRVVAVAKEKNSDLIIGLGGGKTIDSAKAIADKVNLPVVIAPTVASTD 121
V+FNGEASDNEI+RVVA+AKE +D+IIGLGGGKTIDSAKAIAD + +PV+IAPT+ASTD
Sbjct: 61 VAFNGEASDNEISRVAIAKENGNDVIIGLGGGKTIDSAKAIADLLAVPVIIAPTIASTD 120

Query: 122 APTSALSVIYTDEGAPEKYIFYSKNPDLVLVDVTQVIAQAPKRLLASGIADGLATWVEARA 181
APTSALSVIYTDEGAPEKYIFYSKNPDLVLVDVTQVI QAPKRLLASGIADGLATWVEARA
Sbjct: 121 APTSALSVIYTDEGAPEKYIFYSKNPDLVLVDVTQVICQAPKRLLASGIADGLATWVEARA 180

Query: 182 VLQKNGIAMAGGRQTLAGVAIAQACERTLFNDSLQALAACDAKVVTKALENVIEANTLLS 241
V+QKNG MAGG QTLAGVAIA+ACE+TLF D L+A+A+CD +VVT ALENVIEANTLLS
Sbjct: 181 VMQKNGDTMAGGNQTLAGVAIAKACEQTLFADGLKAMASCDRQVVTPALENVIEANTLLS 240

Query: 242 GLGFESAGLAAAHAIHNGFTALSGDIHHLTHGEKVAYGTLTQLFLENRPKEEIDRYINLY 301
GLGFESAGLAAAHAIHNGFTAL+G IHHHLTHGEKVAYGTLTQLFLENR +EEIDRYI+ Y
Sbjct: 241 GLGFESAGLAAAHAIHNGFTALTGAIIHHLTHGEKVAYGTLTQLFLENRSREEIDRYIDFY 300

Query: 302 QAIGMPTTLAELHLGDATYEELKVGQQATTEGETIHEMPFKISAEDVAAALLTVDRYVSN 362
QAIGMPTTL E+HL AT E+ LK+G+QAT+ GETIH+MPF IS EDVAAAL+ VD YV++
Sbjct: 301 QAIGMPTTLKEMHLDATQEDFLKIGRQATMAGETIHQMPFVISPEDVAAALVAVDAYVTS 361

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1112-

Example 1003

A DNA sequence (GBSx1063) was identified in *S.galactiae* <SEQ ID 3079> which encodes the amino acid sequence <SEQ ID 3080>. Analysis of this protein sequence reveals the following:

```

Possible site: 28
5  >>> Seems to have no N-terminal signal sequence
    INTEGRAL    Likelihood = -0.75    Transmembrane  262 - 278 ( 262 - 279)

----- Final Results -----
10      bacterial membrane --- Certainty=0.1298(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

15  >GP:BAA88310 GB:AB028865 O-acetylserine lyase [Streptococcus suis]
    Identities = 239/304 (78%), Positives = 273/304 (89%)

Query: 4   IYNSITDLIGNTPIIQLHHIVPEGAAEVYVKLESFNPSSVKDRIALAMIEDAEQKGLK 63
          IY +IT L+G TP+I+L++IVPEGAAEVYVKLE+FNPGSSVKDRIALAMIEDAE+ G +K
Sbjct: 3   IYQNTITQLVGKTPVIKLNIVPEGAAEVYVKLEAFNPGSSVKDRIALAMIEDAEKAGTIK 62

20  Query: 64  AGDTIVEPTSGNTGIGLAWVGKAKGYNVIIIMPETMSIERRKIIQAYGAQLVLTGSEGM 123
          GDTIVEPTSGNTGIGLAWVG AKGYNVIIIMPETMS+ERRKIIQAYGA+LVLTGSEGM
Sbjct: 63  PGDTIVEPTSGNTGIGLAWVGAAGYNVIIIMPETMSVERRKIIQAYGAELVLTGSEGM 122

25  Query: 124 KGAIKAKEISAEQNAWLPLQFNQANPEIHEKTTGREIIETFGKGLDAFIAGVGTGGT 183
          KGAIKAKEI+ E+N W+P QF N +NP++HE TTG+BI+E FG GLDAF++GVGTGGT
Sbjct: 123 KGAIKAKEIAEEKNGWVPFQFANPSNPKVHEDTTGQRILEDFTGTLDAFVSGVGTGGT 182

30  Query: 184 ITGVSRLKKVNPDVAIYAVEADESAILSGEQPGPHKIQQGISAGFIPETLATDSYDHIIR 243
          ++GVS LK NPD+AIYAVEADESA+LSGE PGPHKIQQGISAGFIP+TL T +YD IIR
Sbjct: 183 VSGVSHVLKTANPDIAIYAVEADESAVLSGEAPGPHKIQQGISAGFIPDTLDT SAYDGIIR 242

35  Query: 244 VTSDDAIETGRIIGLEGFLAGISASAAIYAAIEVAKQLGKGVKVLALLPDNGERYLSTS 303
          V SDDA+ TGR IGG EGFL GIS+ AAI+AAIEVAK+LG GKKVLA+LPDNGERYLST+
Sbjct: 243 VKSDDALATGRAIGGKEGFLVGISSCAAIIHAAIEVAKELGTGKKVLAAILPDNGERYLSTA 302

Query: 304 LYDF 307
          LY+F
Sbjct: 303 LYEF 306
40

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3081> which encodes the amino acid sequence <SEQ ID 3082>. Analysis of this protein sequence reveals the following:

```

Possible site: 58
45  >>> Seems to have no N-terminal signal sequence
    INTEGRAL    Likelihood = -0.48    Transmembrane  262 - 278 ( 262 - 278)

----- Final Results -----
50      bacterial membrane --- Certainty=0.1192(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

55  >GP:BAA88310 GB:AB028865 O-acetylserine lyase [Streptococcus suis]
    Identities = 235/303 (77%), Positives = 261/303 (85%)

Query: 4   IYKTITELVGQTPIIKLNRLIPNEAADVYVKLEAFNPGSSVKDRIALSMIEAAEAEGLIS 63
          IY+ IT+LVG+TP+IKLN ++P AA+VYVKLEAFNPGSSVKDRIAL+MIE AE G I
Sbjct: 3   IYQNTITQLVGKTPVIKLNIVPEGAAEVYVKLEAFNPGSSVKDRIALAMIEDAEKAGTIK 62

60  Query: 64  PGDVIIIEPTSGNTGIGLAWVGAAGYRVIIIMPETMSLERRQIIQAYGAELVLTGPAEGM 123
          PGD I+EPTSGNTGIGLAWVGAAGY VIIIMPETMS+ERR+IIQAYGAELVLTGPA+EGM

```

-1113-

5 Sbjct: 63 PGDTIVEPTSGNTGIGLAWGAAKGYNVIIIVMPETMSVERRKIIQAYGAELVLTTPGSEGM 122
 Query: 124 KGAIAKAETLAIELGAWMPMQFNPNANPSIHEKTTAQEILEAFKEISLDAFVSGVGTGGT 183
 KGAIAKA+ +A E W+P QF NP+NP +HE TT QEILE F LDAFVSGVGTGGT
 10 Sbjct: 123 KGAIAKAKEIAEKNWVPFQFANPSNPKVHEDTTGQEILEDFTTGLDAFVSGVGTGGT 182
 Query: 184 LSGVSHVLKKANPETVIYAVEAEESAVLSGQEPGPHKIQGISAGFIPNTLDTKAYDQIIR 243
 +SGVSHVLK ANP+ IYAVEA+ESAVLSG+ PGPBKIQGISAGFIP+TLDT AYD IIR
 15 Sbjct: 183 VSGVSHVLKTNPDIAIYAVEADESAVLSGEAPGPHKIQGISAGFIPDTLDT SAYDGIIR 242
 Query: 244 VKSKDALETARLTGAKEGFLVGISSGAALYAAIEVAKQLGKGKHLVLTILPDNGERYLSTE 303
 VKS DAL T R G KEGFLVGISSGAA+AAIEVAK+LG GK VL ILPDNGERYLST
 Sbjct: 243 VKSDDALATGRAIGGKEGFLVGISSGAALHAAIEVAKELGTGKKVLAILPDNGERYLSTA 302
 15 Query: 304 LYD 306
 LY+
 Sbjct: 303 LYE 305

An alignment of the GAS and GBS proteins is shown below.

20 Identities = 222/306 (72%), Positives = 263/306 (85%)
 Query: 1 MSKIYNSITDLIGNTPPIQLHHIVPEGAAEVYVKLESFNPSSVKDRIALAMIEDAEQKG 60
 M+KIY +IT+L+G TPII+L+ ++P AA+VYVKLE+FNPSSVKDRIAL+MIE AE +G
 25 Sbjct: 1 MTKIYKTTITELVGQTPIIKLNRLLIPNEAADVYVKLEAFNPGSSVKDRIALSMIEAAEAE 60
 Query: 61 ILKAGDTIVEPTSGNTGIGLAWVGKAKGYNVIIIVMPETMSIERRKIIQAYGAQLVLTTPGS 120
 ++ GD.T+EPTSGNTGIGLAWVG AKGY VIIIVMPETMS+ERR+IIQAYGA+LVLTTPG+
 30 Sbjct: 61 LISPGDVIIIEPTSGNTGIGLAWGAAKGYRVIIIVMPETMSLERRQIIQAYGAELVLTPGA 120
 Query: 121 EGMKGAIKAKEISAEQNAWLPLQFNNOANPEIHEKTTGREIIEFTFGEKGLDAFIAGVGT 180
 EGMKGAIKA+ ++ E AN+P+QFNN ANP IHEKT +EI+E F E LDAF++GVGT
 35 Sbjct: 121 EGMKGAIKAETLAIELGAWMPMQFNPNANPSIHEKTTAQEILEAFKEISLDAFVSGVGT 180
 Query: 181 GGTITGVSRALKKVNPDVAIYAVEADESAILSGEQPGPHKIQGISAGFIPETLATDSYDH 240
 GGT++GVS LKK NP+ IYAVEA+ESA+LSG++PGPHKIQGISAGFIP TL T +YD
 40 Sbjct: 181 GGTLSGVSHVLKKANPETVIYAVEAEESAVLSGQEPGPHKIQGISAGFIPNTLDTKAYDQ 240
 Query: 241 IIRVTSDDAIETGRIIIGGLEGFLAGISASAAIYAAIEVAKQLGKGKVLALLPDNGERYL 300
 IIRV S DA+ET R+ G EGFL GIS+ AA+YAAIEVAKQLGKGK VL +LPDNGERYL
 45 Sbjct: 241 IIRVKSDALETARLTGAKEGFLVGISSGAALYAAIEVAKQLGKGKHLVLTILPDNGERYL 300
 Query: 301 STSLYD 306
 ST LYD
 Sbjct: 301 STELYD 306

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1004

50 A DNA sequence (GBSx1064) was identified in *S.agalactiae* <SEQ ID 3083> which encodes the amino acid sequence <SEQ ID 3084>. Analysis of this protein sequence reveals the following:

Possible site: 56
 >>> Seems to have no N-terminal signal sequence

55 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3666(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

60 >GP:BAB07349 GB:AP001519 unknown conserved protein [Bacillus halodurans]
 Identities = 96/204 (47%), Positives = 127/204 (62%)

-1114-

5 Query: 2 NYKTIKSDGIVEEEIKSRFICHLKRVESEEEGRNYITQIKKAHYKANHSCSAMVIGKEG 61
 +Y T+K GI E I+KSRFI HL R SEEE +I QIKK H+ A H+CSA +IGE
 Sbjct: 4 SYTIVKESGIHEISIQSRFIAHLSRATSEEEAIQFIEQIKKEHWNATHNCSAYLIGEND 63

10 Query: 62 DIKRSSDDGEPSTAGIPMLTVLEKQGLTNVAVVTRYFGGIKLGAGGLIRAYSGSVANT 121
 +++++DDGEPSTAG+PML VL+K+ L + VAVVTRYFGG+KLGAGGLIRAY +V++
 Sbjct: 64 QVQKANDDGEPSGTAGVPMLEVLKKRRLKDTVAVVTRYFGGVKLGAGGLIRAYGSAVSDG 123

10 Query: 122 IKEIGVVEVKEQIGIRIQLTYPQYQTFDNFLKEHHLQEFETEFLEAVTCKIYVDPKEFEH 181
 + IGVVE K I + Y +N L++ H E +LE V + YV E E
 Sbjct: 124 LNAIGVVERKRMQVIHTSIDYHWLGKVENELRQSHYLLKEISYLENVQTYVLEAEVES 183

15 Query: 182 TITNLTEFYQGKALLTEEGSQIVE 205
 +T G+A T + +E
 Sbjct: 184 YCEWMTNLTNGQAAPTHGAIEYLE 207

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3085> which encodes the amino acid sequence <SEQ ID 3086>. Analysis of this protein sequence reveals the following:

20 Possible site: 36
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.43 Transmembrane 86 - 102 (86 - 102)

25 ----- Final Results -----
 bacterial membrane --- Certainty=0.1171(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

30 A related sequence was also identified in GAS <SEQ ID 9153> which encodes the amino acid sequence <SEQ ID 9154>. Analysis of this protein sequence reveals the following:

35 Possible site: 31
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.43 Transmembrane 81 - 97 (81 - 97)

----- Final Results -----
 bacterial membrane --- Certainty=0.1171(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

40 An alignment of the GAS and GBS proteins is shown below.

Identities = 122/206 (59%), Positives = 153/206 (74%)

45 Query: 2 NYKTIKSDGIVEEEIKSRFICHLKRVESEEEGRNYITQIKKAHYKANHSCSAMVIGKEG 61
 ++KTIK+ G EE IKSRFICH+KRV +EE+G+N++ IKK HYKANHSC AM+IG
 Sbjct: 8 HFKTIKASGFFFEESIKSRFICHIKRVSTEEDGKNFVNAIKKEHYKANHSCFAMIIGNNR 67

50 Query: 62 DIKRSSDDGEPSTAGIPMLTVLEKQGLTNVAVVTRYFGGIKLGAGGLIRAYSGSVANT 121
 IKRSDDDGEPSTAGIP+L+VLEKQ LTNVV VVTRYFGGIKLG GGLIRAYS A
 Sbjct: 68 QIKRSDDDGEPSTAGIPILSVLEKQCLTNVAVVTRYFGGIKLGAGGLIRAYSNMTATA 127

55 Query: 122 IKEIGVVEVKEQIGIRIQLTYPQYQTFDNFLKEHHLQEFETEFLEAVTCKIYVDPKEFEH 181
 IK G++EVK+QIG+ I L+YPQYQ + N L + L E ET+F + + +Y D + E+
 Sbjct: 128 IKRFGIIEVKKQIGLEITLSYPQYQLYSNLLDQLALTETETKFSDTIKTTLYCDTERVEN 187

Query: 182 TITNLTEFYQGKALLTEEGSQIVEIP 207
 I LT +Y G+ + GS+++E P
 Sbjct: 188 LIDTLTNYVHGQISCEKIGSKVIEFP 213

60 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1115-

Example 1005

A DNA sequence (GBSx1065) was identified in *S.agalactiae* <SEQ ID 3087> which encodes the amino acid sequence <SEQ ID 3088>. Analysis of this protein sequence reveals the following:

Possible site: 45

5 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

10 bacterial cytoplasm --- Certainty=0.1421(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

15 >GP:AAC44940 GB:U56901 involved in transformation [Bacillus subtilis]
Identities = 160/405 (39%), Positives = 228/405 (55%), Gaps = 20/405 (4%)

Query: 35 YICTRCSSSVAKNCQL----PTGNYYCRECIVFGRVTSNENLYYFPQKTFSTNSLK--W 88
Y C RC + + YCR C++ GRV+ LY + ++ S S+K W
Sbjct: 58 YRCNRCGQTDQRYFSFYHSSGKNKLYCRSCVMMGRVSEEVPLYSWKEENESNWKSIKLTW 117

20 Query: 89 KGELTPYQNEVSEELLKGISSKENLLVHAVTGAGKTEMIYHSVAKVIDTGGSVCIASPRI 148
G+L+ Q + + L++ IS KE LL+ AV GAGKTEM++ + ++ G VCIA+PR
Sbjct: 118 DGKLSGQQAANVLIEALSKKEELLIWAVCGAGKTEMLFPGIESALNQGLRVCIATPRT 177

25 Query: 149 DVCLELYKRLSNDFRCA-ITLMHGESPYSYQR-SPLTIATTHQLLKFYHAFDLLIVDEVDA 206
DV LEL RL F+ A I+ ++G S R SPL I+TTHQLL++ A D++I+DEVDA
Sbjct: 178 DVVLELAPRLKAAPQGADISALYGGSDDKGRLSPLMISTTHQLLRKDAIDVMIIDEVDA 237

Query: 207 FPYVDNPILYQGVKQALKENGTSIFLTATSTTELERKVARKEKLLHLARRFHANPLVIP 266
FPY + L V++A K+N T ++L+AT EL+RK +L + + R H PL P
30 Sbjct: 238 FPYADQTLQFAVQKARKKNSTLVYLSATPPKELKRKALNQGLHSVRIFARHHRKPLPEP 297

Query: 267 EMVWVSGIQKSLQTQKLPPKLYQLINKQRQTRYPLLLFFPHISEGQVFTILRQAFPM EK 326
VW +K L K+PP + + I + P+ LF P +S IL +A K
35 Sbjct: 298 RFVWCGNWKKKLNPNKIPPAVKRWIEFHVKEGRPVFLFVPSVS-----ILEKAAACFK 350

Query: 327 -----IGFVSSKSTSRLLKLVQDFRDNKLSILVSTTILERGVTFPSVDVFIQANHHLFTK 381
V ++ R + VQ FRD +L +L++TILERGV T P V V+ A +FT+
40 Sbjct: 351 GVHCRTASVHAEDKHKRKEKVQQFRDGLDLLITTTILERGVTPVKVQTGVLGAESSIFTE 410

Query: 382 SSLVQISGRVGRALERPEGLLYFLHDGKSKSMHQAIEIKNMNHI 426
S+LVQI+GR GR E +G + + H GK+KSM A K IK MN +
Sbjct: 411 SALVQIAGRTGRHKEYADGDVIYFHFGTKSMLDARKHIKEMNEL 455

45 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3089> which encodes the amino acid sequence <SEQ ID 3090>. Analysis of this protein sequence reveals the following:

Possible site: 21

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -4.09 Transmembrane 304 - 320 (303 - 322)

50 ----- Final Results -----

bacterial membrane --- Certainty=0.2635(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

55 The protein has homology with the following sequences in the databases:

!GB:U56901 involved in transformation [Bacillus subt... 258 1e-67

>GP:AAC44940 GB:U56901 involved in transformation [Bacillus subtilis]

Identities = 155/435 (35%), Positives = 249/435 (56%), Gaps = 20/435 (4%)

60 Query: 10 RLLLESQLPDSAKQLAQPLK-----SVVILRGKMICQRCHYQLDEEA-----RLPSG 56

-1116-

R LL ++L S + + +K S+ I + + C RC Q D+
 Sbjct: 22 RHLLRTELSFSDEMIEWHIKNGYITAENSISINKRRYRCNRCG-QTDQRYFSFYHSSGKN 80
 Query: 57 AYYCRFCLVFGRNQSDKLLYAIPPMHFP--KGNYLWVGGLTAYQEMISQQLLINMONQK 114
 YCR C++ GR + LY+ + K L W G+L++ Q+ + L+ + ++
 Sbjct: 81 KLYCRSCVMGRVSEEVPLYSWKEENESNWKSIKLTWDGKLSGQQAANVLIEAISKKE 140
 Query: 115 TTLVHAVTGAGKTEMIYAAIEAVINTGGWVCIASPRVDVCVEVATRLSQAFS-CSICLMH 173
 L+ AV GAGKTEM++ IE+ +N G VCIA+PR DV +E+A RL AF I ++
 Sbjct: 141 ELLIWAVCAGKTEMLFPFGIESALNQGLRVCIATPRTDVVLELAPRLKAAFQGADISALY 200
 Query: 174 AESLPYQR-APIIVATTHQLLKFKHKAFDLLIIDEVDAPFPVNNIQLHYAASQALKEGGAK 232
 S R +P++++TTHQLL++ A D++IIDEVDAFP+ + L +A +A K+
 Sbjct: 201 GGSDDKGRLSPLMISTTHQLRLYKDAIDVMIIDEVDAPFYSADQTLQFAVQKARKKNSTL 260
 Query: 233 ILLTATSTRTLERKVNKGEVVKLTARRFHNRLVLPKFIKRSFNLFKMIHRQKLPLKILK 292
 + L+AT + L+RK G++ + + R H +PL P+F+ N K ++R K+P + +
 Sbjct: 261 VYLSATPPKELKRKALNGQLHSVRIPARHHRKPLPEPRFVWCGNWKKLNRNKIPPAVKR 320
 Query: 293 YLKKQRKTGYPLLIIFLPTIIMAESVTAILKELPAEQIACVSSSQSNRKEDITAFRQGGK 352
 +++ K G P+ +F+P++ + E A K + + A V ++ ++RKE + FR G+
 Sbjct: 321 WIEFHVKEGRPVFLFVPSVILEKAAACFKGV--HCRTASVHAEDKHKRKEKVQQFRDQGL 378
 Query: 353 TILITTSILERGVTFPQIDVFLGSHHRVYSSQSLVQIAGRVGRSIDRPDGTLYFFHEGI 412
 +LITT+ILERGVT P++ VLG+ +++ +LVQIAGR GR + DG + +FH G
 Sbjct: 379 DLLITTTILERGVTVPKVQTVGLGAESSIFTESALVQIAGRTGRHKEYADGDVIYFHFGEK 438
 Query: 413 SKAMLLARKEIKEMN 427
 +K+ML ARK IKEMN
 Sbjct: 439 TKSMLDARKHIKEMN 453

An alignment of the GAS and GBS proteins is shown below.

Identities = 223/427 (52%), Positives = 299/427 (69%)

Query: 1 MENYLGRWLTKAQLSEQLRKIAISLPSFIKKGSDYICTRCSSSVAKNCQLPTGNYYCREC 60
 +EN GRL ++QL + +++A L S + IC RC + + +LP+G YYCR C
 Sbjct: 4 IENSYGRLLLESQLPDSAKQLAQPLKSVILRGKMICQCHYQLDEEARLPSGAYYCRFC 63
 Query: 61 IVFGRVTSNENLYFFPQKTFSTNSLKWKGELTPYQNEVSEELLKGISSKENLLVHAVTG 120
 +VFGR S++ LY P F K N L W G+LT YQ +S++LL + +++ LVHAVTG
 Sbjct: 64 LVFGRNQSDKLLYAIPPMHFPKGNYLWVGGLTAYQEMISQQLLINMONQKTTLVHAVTG 123
 Query: 121 AGKTEMIYHSVAKVIDTGGSVCIASPRIDVCLLEYKRLSNDFRCAITLMHGESPYSYQSRP 180
 AGKTEMIY ++ VI+TGG VCIASPR+DVC+E+ RLS F C+I LMH ES YQR+P
 Sbjct: 124 AGKTEMIYAAIEAVINTGGWVCIASPRVDVCVEVATRLSQAFSCSICLMHAESLPYQRAP 183
 Query: 181 LTIATTHQLLKFYHAFDILLIVDEVDAPFPVDNPIYQGVKQALKENGTSIFLTATSTTEL 240
 + +ATTHQLLK+ AFDLLI+DEVDAFP+V+N L+ QALKE G I LTATST L
 Sbjct: 184 IIVATTHQLLKFKHKAFDLLIIDEVDAPFPVNNIQLHYAASQALKEGGAKILLTATSTRTL 243
 Query: 241 ERKVARKELKKLHLARRFHANPLVIPMEMVWVSGIQSLQTQKLPPKLYQLINKQRQTRY 300
 ERKV + E+ KL LARRFH PLVIP+ + + K + QKLP K+ + + QQR+T YP
 Sbjct: 244 ERKVNGGEVVKLTARRFHNRLVLPKFIKRSFNLFKMIHRQKLPLKILKYLKQKRTGY 303
 Query: 301 LLLFFPHISEGQVFTEILRQAFPMKIGFVSSKSTSRKLKVQDFRDNKLSILVSTTILER 360
 LL+F P I + T IL++ P E+I VSS+S +R + + FR K +IL++T+ILR
 Sbjct: 304 LLIFLPTIIMAESVTAILKELPAEQIACVSSSQSNRKEDITAFRQGGKKTILITTSILER 363
 Query: 361 GVTFPVDVVFVIQANHHLFTKSSLVQISGRVGRALERPEGLLYFLHDGKSKSMHQAIKEI 420
 GVTFP +DVFV+ ++H +++ SLVQI+GRVGR+++RP+G LYF H+G SK+M A KEI
 Sbjct: 364 GVTFPQIDVFLGSHHRVYSSQSLVQIAGRVGRSIDRPDGTLYFFHEGISKAMLLARKEI 423
 Query: 421 KNNMHIG 427
 K MN+ G
 Sbjct: 424 KEMNYKG 430

-1117-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1006

A DNA sequence (GBSx1066) was identified in *S.agalactiae* <SEQ ID 3091> which encodes the amino acid sequence <SEQ ID 3092>. This protein is predicted to be comf operon protein 3 (comFC). Analysis of this protein sequence reveals the following:

Possible site: 41
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.0894 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC44942 GB:U56901 involved in transformation [Bacillus subtilis]
Identities = 76/230 (33%), Positives = 118/230 (51%), Gaps = 11/230 (4%)

Query: 1 MTCLLCHEIDLSQLTFVELMLLPKQNVICQTCKGSFEALSREMGCQTCK-QIPQKQCQ 59
M CLLC +T+ L LLKP + V C +C+ + ++ + C C + Q C+
Sbjct: 1 MICLLCDSQFSQDVTWRALFLLKPDEKV-CYSCRSKLKKTIGHI-CPLCGRPQSVHAVCR 58

Query: 60 DCIYWGKKGIEV----NHFSLYRYNEAMKKNFSLFKFGDYLLKDVFTKEIKAALKKY-- 113
DC W + + + S+Y YN+ MK+ S FKF+GD + + F + + K
Sbjct: 59 DCEVWRTRIRDSLLLRQNRSVYTYNDMMKETLSRFKFRGDAEIIINAFKSDFSSTFSKVYP 118

Query: 114 -KGYTIVPVPLSHEGYQNRQFNQVIAFLQSANIPYKNILSKKDGKQSANNKEERLKQVQ 172
K + +VP+PLS E + R FNQ + + P + L + + QQS K ERL
Sbjct: 119 DKHFVLVPIPLSKEREERGERGFAHLLAECLDRPSHHPLIRLNNEKQSKKKKTERLLSEC 178

Query: 173 QFTLKNEAELGDNLLIVDDIYTTGATIAQIRKLEEKG-IKNIKSFSLAR 221
F KN + G N+++DD+YTTGAT+ + L EKG ++ SF+L R
Sbjct: 179 IFDTKNNSAEGMNIILIDDLTYTTGATLHFAARCLLEKGAASVSSFTLIR 228

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3093> which encodes the amino acid sequence <SEQ ID 3094>. Analysis of this protein sequence reveals the following:

Possible site: 52
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.0763 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 100/222 (45%), Positives = 139/222 (62%), Gaps = 2/222 (0%)

Query: 1 MTCLLCHEIDLSQLTFVELMLLPKQNVICQTCKGSFEALSREMGCQTCKQIPQKQCQD 60
M CLLC +I + ++ E++ L+ + ICQ C+ SF+ + + + C TCC C+D
Sbjct: 1 MICLLCQQISQTPISTITEIIFLRRISSPICQQCKSFQKIGKSV-CATCCANSDIACRD 59

Query: 61 CIYWGKKGIEVNHFSLYRYNEAMKKNFSLFKFGDYLLKDVFTKEIKAALKKY-KGYTIV 119
C+ W KG VNH SLY YN AMK FS +KFQGDYLL+ VF E+ + KY KGY V
Sbjct: 60 CLKWENKGVNVNHRSLYCYNAAAMKAYFSQYKFGDYLLRKVFAVELADVITKYVKGYPV 119

Query: 120 PVPLSHEGYQNRQFNQVIAFLQSANIPYKNILSKKDGKQSANNKEERLKQVQQFTLKNE 179
PVP+S ++ RQFNQV A L++AN+ Y ++ K D QS+ K+ERL + + L
Sbjct: 120 PVPVSPGCFRERQFNQVSAILEAANVSYSLEKLDNTHQSSRTKKERLLVEKSYRLLKV 179

-1118-

Query: 180 AELGDNLLIVDDIYTTGATIAQIRKLLLEEKGIKNIKSFSLAR 221
 + + D +LIVDDIYTTG+TI +RK L + +IKS S+AR
 Sbjct: 180 SNIPDKILIVDDIYTTGSTIIALRKQLAKVANSDIKSLSIAR 221

5

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1007

A DNA sequence (GBSx1067) was identified in *S.agalactiae* <SEQ ID 3095> which encodes the amino acid sequence <SEQ ID 3096>. Analysis of this protein sequence reveals the following:

Possible site: 16
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

15 bacterial cytoplasm --- Certainty=0.3889(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

20 >GP:CAB91549 GB:AJ249134 hypothetical protein [Lactococcus lactis]
 Identities = 107/185 (57%), Positives = 140/185 (74%), Gaps = 3/185 (1%)

Query: 1 MIKYSIRGENIEVTEAIREYVETKLSKVEKYFNEAQELDTRVNLKVYREKTAKVEVTILI 60
 MIK++IRGEN+EVT+AIR YVE K+ K++KYFN+ E+ VNLKVY EK AKVEVT+

25 Sbjct: 1 MIKFNIRGENVEVTDARAYVEDKIGKLDKYFNDGHEVTAYVNLKVYTEKRAKVEVTLP 60

Query: 61 DSITLRAEDVSQDMYGSIDLVDKIERQIRKNKTKIAKKYREKIPASQVFTTEFEAEPE 120
 ++TLRAED SQDMY SID V +K+ERQIRK KT++ +K R +P QVF EF

30 Sbjct: 61 KNVTLRAEDTSQDMYSSIDFVEEKLERQIRKYKTRMNRKPRNAVPTGQVFGDEFAPLDTT 120

Query: 121 EAVSQ---RIVRTKNVNLKPMDEEALLQMEELGHDFFIYTDADNTTNVLYKREDGELG 177
 + V++ IVRTK+V LKPMDEEA+LQM++LGHDF+++TDA+ N T+V+Y+R DG G

35 Sbjct: 121 DEVAEDHVDIVRTKHVALKPMDEEAVALQMDMLGHDFYVFTDADSNNGTHVVYRRTDGRYG 180

Query: 178 LIEAK 182
 LIE +

Sbjct: 181 LIETE 185

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3097> which encodes the amino acid sequence <SEQ ID 3098>. Analysis of this protein sequence reveals the following:

Possible site: 16
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

45 bacterial cytoplasm --- Certainty=0.3751(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

50 Identities = 145/182 (79%), Positives = 165/182 (89%)

Query: 1 MIKYSIRGENIEVTEAIREYVETKLSKVEKYFNEAQELDTRVNLKVYREKTAKVEVTILI 60
 MIK+SIRGENIEVTEAIR+YVE+KL+K+EKYF + QE+D RVNLKVYRE+++KVEVTI +

55 Sbjct: 1 MIKFSIRGENIEVTEAIRDYVESKLTKIEKYFAKDQEIDARVNLKVYRERSKVEVTIPL 60

Query: 61 DSITLRAEDVSQDMYGSIDLVDKIERQIRKNKTKIAKKYREKIPASQVFTTEFEAEPE 120
 DS+TLRAEDVSQDMYGSIDLVDKIERQIRKNKTKIAKK+REK+P QVFTTEFEAE +

Sbjct: 61 DSVTLRAEDVSQDMYGSIDLVDKIERQIRKNKTKIAKKHREKVPTGQVFTTEFEAEVD 120

-1119-

Query: 121 EAVSQRIVRTKNVNLKPMDEV EALLQMELLGHDFFIYTD AEDNTTNVLYKREDGELGLIE 180
 E ++VRTKNV LKPMDEV EEA LQMELLGHDFFIYTD+ED TN+LY+REDG LGLIE
 Sbjct: 121 EIPEVQVVRTKNVTLKPMDEV EEARLQMELLGHDFFIYTD SEDGATNILYRREDGNLGLIE 180

5 Query: 181 AK 182
 AK
 Sbjct: 181 AK 182

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 10 vaccines or diagnostics.

Example 1008

A DNA sequence (GBSx1068) was identified in *S.agalactiae* <SEQ ID 3099> which encodes the amino acid sequence <SEQ ID 3100>. Analysis of this protein sequence reveals the following:

15 Possible site: 16
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0685(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 20 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for
 25 vaccines or diagnostics.

Example 1009

A DNA sequence (GBSx1077) was identified in *S.agalactiae* <SEQ ID 3101> which encodes the amino acid sequence <SEQ ID 3102> (sgaT). Analysis of this protein sequence reveals the following:

30 Possible site: 41
 >>> Seems to have a cleavable N-term signal seq.

INTEGRAL	Likelihood = -5.95	Transmembrane	99 - 115 (87 - 115)
INTEGRAL	Likelihood = -3.50	Transmembrane	43 - 59 (42 - 60)

----- Final Results -----
 35 bacterial membrane --- Certainty=0.3378(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

40 >GP:BAB03942 GB:AP001507 unknown conserved protein [Bacillus halodurans]
 Identities = 47/111 (42%), Positives = 76/111 (68%), Gaps = 5/111 (4%)

Query: 1 MAIIYLIVAVFAG--EAYIAKEI---SNGVNGLVYALQLAGQFAAGVFVILAGVRLIIGE 55
 M I++L+ A+ + A+E+ S + +YA+ + FA G+ V+L GV++ +GE
 45 Sbjct: 233 MGILFLVGAIILALKDTQGAQELIAQSGEQSFFIYAI IQSFMFAGGIAVVLGVKMFIFE 292

Query: 56 IVPAPKGI SEKLV P NSK PALDCPIVVPYAPNAVLIGFISK FVGGLVSMIVM 106
 +VPAF GI+ KLVP ++PALD P+V+P APNAV++GF+ FVG L+ ++V+
 50 Sbjct: 293 VVPAFNGIATKLVPGARPALDAPVVFFMAPNAVILGFLGAFVGALIVLVVI 343

There is also homology to SEQ ID 516.

-1120-

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1010

A DNA sequence (GBSx1078) was identified in *S.agalactiae* <SEQ ID 3103> which encodes the amino acid sequence <SEQ ID 3104>. This protein is predicted to be tryptophanyl-tRNA synthetase (trpS). Analysis of this protein sequence reveals the following:

Possible site: 50
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2156(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAC05711 GB:L49336 tryptophanyl-tRNA synthetase [Clostridium longisporum]
Identities = 225/340 (66%), Positives = 271/340 (79%), Gaps = 3/340 (0%)

Query: 1 MTKPIILTGDRPTGKLHGHYVGS�KKNRVLLQNEGSYTLFVFLADQQALTDHAKDPQTIV 60
M K IILTGDRPTGKLHGHYVGS�KKNRV LQN G Y F+ +ADQQALTD+A++P+ I
Sbjct: 1 MAKEIILTGDRPTGKLHGHYVGS�KKNRVQLQNSG DYRSFIMIADQQALTDNARNPEKIR 60

Query: 61 ESIGNVALDYLA VGLDPNKSTLFIQSQIPELAELSMYMNVLVSLARLERNPTVKTEIAQK 120
S+ VALDYLA VGD+DP KST+ +QSQIPEL EL+M+Y+NLV+L+RLERNPTVK EI QK
Sbjct: 61 NSLIEVALDYLA VGDIDPLKSTILVQSQIPELNLTMHYLNLVTLRRLERNPTVKA EIKQK 120

Query: 121 GFGESIPAGFLVYPVAQAADITAFKANLVFVGTDQKPMIEQTR EIVRSFNHAYNCQVLVE 180
F SIPAGFL+YPV+QAADITAFKA VPVG DQ PMIEQ REIVRSFN Y +VLVE
Sbjct: 121 NFENSIPAGFLIYPVQAADITAFKATTVPVGEDQLPMIEQAREIVRSFNTIYGKEVLVE 180

Query: 181 PEGIYPENDAGRLPGLDGNAKMSKSLNNGIFLADDMDTVKKVMSMYTDPNHIKVEEPG 240
P+ + P+ GRLPG DG AKMSKS+ N I+LAD+ D +K+KVMSMYTDPNHIKV +PG
Sbjct: 181 PKAVIPKG-TIGRLPGTDGKAKMSKSIGNAIYLADEADV I KQKVMSTDPNHIKVTDPG 239

Query: 241 QIEGNMVFHYLDVFGREDEQKEITAMKEHYQGGGLGDVKT KRYLLDILERELSPIRERRL 300
Q+EGN VF YLD F +D + E MK HY +GGLGDVK K+L +IL+ EL PIR RR
Sbjct: 240 QVEGNTVFTYLDTFCKDTETLE--EMKAHYSRGGLGDVKKVFLNEILQAELEPIRNRK 297

Query: 301 EYAKDMGQVYQMLQKGSEKAQAVAASTLDEVKSAMGLNYF 340
E+ KD+ +VY++L++GSEKA+ VAA TL EV+ +G+ YF
Sbjct: 298 EFQKDIPEVYRILKEGSEKAREVAAGTLKEVRETIGIEYF 337

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3105> which encodes the amino acid sequence <SEQ ID 3106>. Analysis of this protein sequence reveals the following:

Possible site: 54
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2737(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 290/340 (85%), Positives = 316/340 (92%)

Query: 1 MTKPIILTGDRPTGKLHGHYVGS�KKNRVLLQNEGSYTLFVFLADQQALTDHAKDPQTIV 60
MTKPIILTGDRPTGKLH+GHYVGS�KKNRV LQNE Y +FVFLADQQALTDHAK+ + I

-1121-

Sbjct: 2 MTKPIILTGRPTGKHLGHVYVGSILQNRVFLQENKYKMFVFLADQQALTDHAKESLIQ 61

Query: 61 ESIGNVALDYLAVALDPNKSTLFIQSQIPELAELSMYYMNLVSLARLERNPTVKTEIAQK 120
ESIGNVALDYLAVALDP +ST+FIQSQIPELAELSMYYMNLVSLARLERNPTVKTEIAQK

5 Sbjct: 62 ESIGNVALDYLSVGLDPKQSTIFIQSQIPELAELSMYYMNLVSLARLERNPTVKTEIAQK 121

Query: 121 GFGESIPAGFLVYPVAQAADITAFKANLVPVGTQKPMIEQTREIVRSFNHAYNCQVLVE 180
GFGESIP+GFLVYPV+QAADITAFKANLVPVG DQKPMIEQTREIVRSFNH Y+ LVE

10 Sbjct: 122 GFGESIPSGFLVYPVVSQAADITAFKANLVPVGNQKPMIEQTREIVRSFNHTYHTDCLVE 181

Query: 181 PEGIYPENDAAGRLPGLDGNAMSKSLNNGIFLADDMDTVKKVMSMYTDPNHKIVEEPG 240
PEGIYPEN+ AGRLPGLDGNAMSKSL NGI+L+DD DTV+KKVMSMYTDPNHK+E+PG

Sbjct: 182 PEGIYPENKAGRLPGLDGNAMSKSLGNGIYLSDDADTVRKKVMSMYTDPNHKIEDPG 241

15 Query: 241 QIEGNMVFHYLDVFRDEQKEITAMKEHYQGGGLGDVTKRYLLDILERELESPRERRL 300
QIEGNMVFHYLD+F R EDQ +I AMKEHYQ GGLGDVTKRYLLDILERELE+PIRERRL

Sbjct: 242 QIEGNMVFHYLDIFARKEDQADIEAMKEHYQIGGLGDVTKRYLLDILERELEPIRERRL 301

Query: 301 EYAKDMGQVYQMLQKGSEKAQAAVASTLDEVKSAMGLNYF 340
EYAKDMG+V++MLQ+GS+KA+ VAA TL EVKSAMG+NYF

20 Sbjct: 302 EYAKDMGEVFRMLQEGSQKARTVAAKTLSEVKSAMGINYF 341

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

25 Example 1011

A DNA sequence (GBSx1079) was identified in *S.agalactiae* <SEQ ID 3107> which encodes the amino acid sequence <SEQ ID 3108>. This protein is predicted to be carbamate kinase. Analysis of this protein sequence reveals the following:

Possible site: 24
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

30 bacterial cytoplasm --- Certainty=0.0013(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
35 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA04684 GB:AJ001330 carbamate kinase [Lactobacillus sakei]
Identities = 199/311 (63%), Positives = 254/311 (80%), Gaps = 3/311 (0%)

40 Query: 6 QKIVVALGGNAILSTDASAKAQEALINTSKSLVKLIKEGHDVIVTHGNGPQVGNLLLLQQ 65
+KIVVALGGNAILSTDASA AQ +A+ T K LV +K+G +I++HGNGPQVGNLL+QQ

Sbjct: 4 RKIVVALGGNAILSTDASANAQIKAVKETVKQLVAFVKQGDQLIISHGNGPQVGNLLLIQQ 63

45 Query: 66 AASDSEKNPAMPLDTCVAMTEGSIGFWLQNALNNELQEQQIDKEVATVVTQVIVDEKDQA 125
AASDSEK PAMPLDT AM++G IG+W+QNA N L E+G+ +VAT+VTQ IVD KD+A

Sbjct: 64 AASDSEKTPAMPLDTVGAMSQGEIGYWMQNAFNEVLAEEGLALDVATIVTQTIVDAKDEA 123

50 Query: 126 FTNPTKPIGPFLSEDAKKQAQ-ETGSKFKEDAGRGWRKVPSPKPVGIKEASVIRRLVD 184
F NPTKPIGPF SE +AKKQ + F EDAGRGWR+VVPSP+P+GI+EA VI++LV+

Sbjct: 124 FQNPTKPIGPFIYSEAEAKKQQSINPEAHFVEDAGRGWRRVVPSPRPIGIQEAPVIQKLVE 183

Query: 185 SGVVVISAGGGGVPIEDANTKALKGVEAVIDKDFASQTLSELVDADLFIVLTGVDNVFV 244
V+ ISAGGGGVPV ++ N L+GVEAVIDKDFAS+ L+ELV AD+ I+LT VDNV+V

55 Sbjct: 184 GNVITISAGGGGVPAKEGN--KLRGVEAVIDKDFASEKLAELVGADMLIILTAVDNVYV 241

Query: 245 NFNKPNQEKLEEVTVSQMKQYITENQFAPGSMLPKVEAAIAFVENKPESRAIITSLENID 304
NFNKP+Q+KL V+V++++ YI ++QFA GSMLPK++ AI +V N+P+S+AIITSL+N+

60 Sbjct: 242 NFNKPDQKCLTNVSAELEDYIKDDQFAKGSMLPKIQTAIYVNNRPDSKAIITSLDNVK 301

Query: 305 NVLAQNAGTQI 315

-1122-

N+LA +AGT I
Sbjct: 302 NLLAHDAGTII 312

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3109> which encodes the amino acid
5 sequence <SEQ ID 3110>. Analysis of this protein sequence reveals the following:

Possible site: 22
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
10 bacterial cytoplasm --- Certainty=0.0013 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 275/312 (88%), Positives = 295/312 (94%)
Query: 6 QKIVVALGGNAILSTDASAKAQEALINTSKSLVKLIKEGHVIVTHGNGPQVGNLLLQQ 65
QKIVVALGGNAILSTDASAKAQEALI+TSKSLVKLIKEGH+VIVTHGNGPQVGNLLLQQ
Sbjct: 4 QKIVVALGGNAILSTDASAKAQEALISTSKSLVKLIKEGHEVIVTHGNGPQVGNLLLQQ 63
20 Query: 66 AADSEKNPAMPLDTCVAMTEGSIGFWLQNALNNEQEGIDKEVATVVTQVIVDEKDQA 125
AA+DSEKNPAMPLDTCVAMTEGSIGFWL NAL+NELQ QGI KEVA VVTQVIVD KD A
Sbjct: 64 AADSEKNPAMPLDTCVAMTEGSIGFWLVNALDNELQAQGIQKEVAAVVTQVIVDAKDPA 123
25 Query: 126 FTNPTKPIGPFLLSEEDAKKQAQETGSKFKEDAGRGWRKVVPSPKPVGIKEASVIRRLVDS 185
F NPTKPIGPFLL+EEDAKKQ E+G+ FKEDAGRGWRKVVPSPKPVGIKEA+VIR LVDS
Sbjct: 124 FENPTKPIGPFLLTEEDAKKQMAESGASFKEAGRGWRKVVPSPKPVGIKEANVIRSLVDS 183
30 Query: 186 GVVVISAGGGGVPVIEDANTKALKGVEAVIDKDFASQTLSELVDADLFIVLTGVDNVFVN 245
GVVV+SAGGGGVPV+EDA +K L GVEAVIDKDFASQTLSELVDADLFIVLTGVDNV+VN
Sbjct: 184 GVVVISAGGGGVPVVEDATSKTLTGVEAVIDKDFASQTLSELVDADLFIVLTGVDNVYVN 243
35 Query: 246 FNKPNQEKLEEVTVSQMKQYITENQFAPGSMLPKVEAAIAFVENKPESRAIITSLENIDN 305
FNKP+Q KLEEVTVSQMK+YIT++QFAPGSMLPKVEAAIAFVENKP ++AIITSLENIDN
Sbjct: 244 FNKPDQAKLEEVTVSQMKYIITQDQFAPGSMLPKVEAAIAFVENKPNAKAIITSLENIDN 303
40 Query: 306 VLAQNAGTQIVA 317
VL+ NAGTQI+A
Sbjct: 304 VLSANAGTQIIA 315

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1012

A DNA sequence (GBSx1080) was identified in *S.agalactiae* <SEQ ID 3111> which encodes the amino
45 acid sequence <SEQ ID 3112>. This protein is predicted to be permease (potE). Analysis of this protein sequence reveals the following:

Possible site: 52
>>> Seems to have an uncleavable N-term signal seq

50 INTEGRAL Likelihood = -12.63 Transmembrane 450 - 466 (441 - 478)
INTEGRAL Likelihood = -8.97 Transmembrane 236 - 252 (231 - 259)
INTEGRAL Likelihood = -8.70 Transmembrane 283 - 299 (277 - 308)
INTEGRAL Likelihood = -8.44 Transmembrane 165 - 181 (153 - 186)
INTEGRAL Likelihood = -7.96 Transmembrane 129 - 145 (126 - 151)
55 INTEGRAL Likelihood = -6.16 Transmembrane 396 - 412 (394 - 415)
INTEGRAL Likelihood = -5.15 Transmembrane 45 - 61 (38 - 63)
INTEGRAL Likelihood = -4.94 Transmembrane 335 - 351 (334 - 352)
INTEGRAL Likelihood = -3.72 Transmembrane 13 - 29 (10 - 30)
INTEGRAL Likelihood = -2.92 Transmembrane 417 - 433 (417 - 435)
INTEGRAL Likelihood = -1.54 Transmembrane 360 - 376 (360 - 376)

-1123-

INTEGRAL Likelihood = -0.53 Transmembrane 207 - 223 (207 - 223)

----- Final Results -----

5 bacterial membrane --- Certainty=0.6052(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10295> which encodes amino acid sequence <SEQ ID 10296> was also identified.

10 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA76779 GB:Y17554 permease [Bacillus licheniformis]

Identities = 265/470 (56%), Positives = 347/470 (73%), Gaps = 3/470 (0%)

15 Query: 5 MEKEKKLGLPLTMLVIGSLIGGGIFDLMQNMSSRAGLVPMLIAWVITAIGMGTFVLSFQ 64
M +EKKLGL L LVIGS+IGGG F+L +M+S AG +LI W+IT +GM SFQ
Sbjct: 1 MAEKKLGLFALIALVIGSMIGGGAFNLASDMASGAGAGAILIGWIITGVGMIALAFSFQ 60

20 Query: 65 NLSEKRPDLTAGIFSYAKEGFGNFMGFNSAWGYWLSAWLGNVAYAAALLFSSLGYYFFKFFG 124
NL+ KRPD L GIF+YA+EGFG+FMGFNS WGYW +A LGNVAY LLFS++GYF FG
Sbjct: 61 NLITKRPDLDGGIFTYAREGFGHFMGFNSGWGYWFAALLGNVAYGTLLFSAIGYFIPAFG 120

25 Query: 125 NGNNIISIIGASIVIVVHFLILRGVNTAAFINITIVTFAKLVPVLIIFLISALLAFKFNIF 184
+G NI SIIGAS+++W VHFILILRGV +AA IN I T +KLVP+ F+I+ + F ++F
Sbjct: 121 DGQNIASIIGASVILWCVHFLILRGVQSAAMINLITISKLVPIFAFIIAIFVFHLDLF 180

30 Query: 185 SLDIWGNGLH-QSIFNQVNSTMTAVVWVFIGIEGAVVFSGRAKKHSIDIGKASILALFTMI 243
+ D WG GL SI QV STM VVWF GIEGAV+FS RAKK SD+GKA+++ L +++
Sbjct: 181 TNDPFWGKLSLSIGTQVKSTMLVTVWVFTGIEGAVLFSSRAKSSDVGKATVIGLISVL 240

35 Query: 244 SLYVLISVLSLIGMSRPELANLKTTPAMAYVLEKAVGHWGAILVNLGVIIISVFGAILAWTL 303
+YV+I++LSLG+M++ LA L P+MA ++E VG WGA+L+NLG+IISV GA LAWTL
Sbjct: 241 VIYVMITMLSLGVMNQNLAEPLNPSMAAIMEHIVGKWGAVINLGLIISVLGAWLAWTL 300

40 Query: 304 FAELPYQAAKEGAFPKFFAKENKKNKAPINSLVTNLCVQAPLITFLFTQSAYRFGFALA 363
FA ELP AA+EG FPK+F KENKN AP N+L +TN +Q FL+TFL + +AY+F F+LA
Sbjct: 301 FAGELPLIAAREGVFPKWFGENKNGAPTALTLTNAILQLFLTLFLISDAAYQFAFSLA 360

45 Query: 364 SSAILIPYAFTALYQLQFTLREDKSTPGHQKNLIIGILATIYAVYLIYAGGFVYLLLTMI 423
SSAILIPY F+ LYQL+++ + F KNLIIGI+A+IY V+L+YA G DYLLLTMI
Sbjct: 361 SSAILIPYLFSGLYQLKYSLWHLKE--PNRGKNLIIGIASIYGVWLVIYAAGLDYLLLTMI 418

50 Query: 424 AYTGLMILYIKMRKDDKLPFVGYEKISAIVILALCLLCIIEIMTGQIDI 473
Y G++++ +RK + P+F E + A +IL L ++ +I + +G I I
Sbjct: 419 LYAPGILVFRVRKKGKGPVFNKAELLIALILVAVIAVIRLASGSISI 468

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3113> which encodes the amino acid sequence <SEQ ID 3114>. Analysis of this protein sequence reveals the following:

Possible site: 51

>>> Seems to have no N-terminal signal sequence

50 INTEGRAL Likelihood = -11.52 Transmembrane 331 - 347 (327 - 354)
INTEGRAL Likelihood = -9.50 Transmembrane 390 - 406 (383 - 410)
INTEGRAL Likelihood = -8.12 Transmembrane 50 - 66 (45 - 75)
INTEGRAL Likelihood = -7.59 Transmembrane 235 - 251 (234 - 262)
INTEGRAL Likelihood = -6.21 Transmembrane 133 - 149 (128 - 151)
55 INTEGRAL Likelihood = -5.84 Transmembrane 162 - 178 (153 - 183)
INTEGRAL Likelihood = -2.02 Transmembrane 105 - 121 (105 - 121)
INTEGRAL Likelihood = -1.49 Transmembrane 414 - 430 (414 - 431)
INTEGRAL Likelihood = -0.69 Transmembrane 280 - 296 (280 - 296)
60 INTEGRAL Likelihood = -0.59 Transmembrane 21 - 37 (21 - 37)
INTEGRAL Likelihood = -0.32 Transmembrane 205 - 221 (205 - 222)

----- Final Results -----

bacterial membrane --- Certainty=0.5607(Affirmative) < succ>

-1124-

bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

5 >GP:AAB85052 GB:AE000837 cationic amino acid transporter related
 protein [Methanobacterium thermoautotrophicum]
 Identities = 108/422 (25%), Positives = 213/422 (49%), Gaps = 36/422 (8%)

10 Query: 26 INAVIGSGIFLLPRAIYKGLGPASIAVMFGTAILTITMLAVCFAEVSGYFGKNGGAFQYSK 85
 + ++G+ I+++ LGPASI ++ +++A+ F+E S + GG + Y+
 Sbjct: 19 VGTIVGADIYIVAAYGAGSLGPASILAWLLAGLMALIILVFSEASAMLPRTGGPYVYAG 78

Query: 86 RAFGDFIGFNVGFLGWTVTIFAWAAMAAGFARMFIITPFAFEGWHIPL-----SIGL 137
 A G F GF GW++ + +W A+A +F + F + + IPL +
 15 Sbjct: 79 EALGRFTGF---ITGWSLWVSSWVAIA-----VPPLAFIYXLEYFIPLDPPAEAVIKVLF 130

Query: 138 IILLSLMNIAGLKTSKIIVTITATIAKLIPIVAFCACTLFFIKNG----LPNFTPFVQLEP 193
 I+ L+++NIAG+ + V TI K+ P++ F + + N+TP +
 20 Sbjct: 131 ILSLTIINIAGVGRAGKVNDILTILKVAPVLLFAVLGAIHLALNPGLLVSNYTPAAPMG- 189

Query: 194 GTNLLGAISNTAVYIFYGFIGFETLSIVAGEMRDPKENVPRALLGSISIVSVLYMLIIGG 253
 LGA+ V +F+ ++GFE +++ A E+RDPE+ +P ++ + V++ Y+L
 Sbjct: 190 ----LGALGTVTVLVFWAYVGFELVTPADEVRDPERTIPLSITLGMIFVTLFYILTNAV 245

25 Query: 254 TIAMLGSIQIMMTN-APVQDAFVKMIGPAGAWMVSIKALISITGLNMGESIMVPRYGAAIA 312
 + ++ +++ ++ AP+ A ++G GA +++ GA+ SI G + R A++
 Sbjct: 246 ILGLVFWRVLASSTAPLTVAGYSLMGGIGALILTAGAVFSIAGSEEGMLTARLLFAMS 305

30 Query: 313 DEGLLPAAIAKQONQON-GAPLVAILVSGAIAIVLLLTGSPFESLAKLSVVFRFQYIPTALA 371
 ++G LP +++ ++ G P ++ILV A++ LTG+ L +LSVV Y T ++
 Sbjct: 306 EDGFLPGFLSRVHRRFGTTPHMSILVQNLTAALTAGTVSGLIELSVVTLTLLPYAVTCIS 365

Query: 372 VMKLRKDDPDANVIFRVFPGPIIPILAVIVSLVMIWGDNPMPNFVYGAAGVVIASSVYYLM 431
 + LR+ D P+ +L V+V + ++ P +G + +I++ + YL+
 35 Sbjct: 366 LAILRRRDGSGI-----PLKSVLGVLCIYLLMNTTTPSTTAWGLL-LILSGAPLYLI 416

Query: 432 HG 433
 G
 40 Sbjct: 417 FG 418

An alignment of the GAS and GBS proteins is shown below.

Identities = 104/368 (28%), Positives = 162/368 (43%), Gaps = 32/368 (8%)

45 Query: 1 MRYKMEKEKIKLGLPLTMLVIGSLIGGGIFDLMQNSSRAGLVPMIAWVI-TAIGMGTF 59
 M + ++ K L T+ I ++IG GIF L + + GL P IA + TAI
 Sbjct: 6 MNEQBEREQAKFSLSCATLYGINAVIGSGIFLLPRAIYK--GLGPASIAVMFGTAILTITML 63

Query: 60 VLSFQNLSEKRPDLTAGIFSYAKEGFGNFMGFNSA--WGYWLSAWLGNVAYAAALLFSSL 116
 + F +S G F Y+K FG+F+GFN W + AW A A +F
 50 Sbjct: 64 AVCFAEVSGYFGK-NGGAFQYSKRAFQDFIGFNVGFLGWTVTIFAWAAMAAGFARMFIIT 122

Query: 117 GYFFKFFGNGNNIISIGASIVIVVHFLILRGVNTAAFINTIVTFKLVPIIPLISAL 176
 F+ G +I IG I++ +++ + G+ T+ + T AKL+P++ F L
 55 Sbjct: 123 FPAFE----GWHIPLSIGLIILLSLMN---IAGLKTSKIIVTITATIAKLIPIVAFCACTL 175

Query: 177 LAFK-----FNIFSLDIWGNGLHQSIQNVNSTMKTAVVVFIEGAVVFSGRAKKHSI 231
 K F F G L +I N TAV++F G G S A + D
 60 Sbjct: 176 FFIKNGLPNFTPFVQLEPGTNLLGAISN-----TAVYIFYGFIGFETLSIVAGEMRDP 228

Query: 232 GKASILALFTMISLYVLISVLSLG---IMSRPELANLKTAPAM-AYVLEKAVGHWGAILVN 287
 K AL IS+ ++ +L +G M ++ P A+V K +G GA +V+
 65 Sbjct: 229 EKNVPRALLGSISIVSVLYMLIIGGTIAMLGSIQIMMTNAPVQDAFV--KMIGPAGAWMV 286

Query: 288 LGVIIISVFGAILAWTLFAELPYQAARKGAFFPKFFAKENKNKAPINSLLVTNLCVQAFLI 347
 +G +IS+ G + ++ A EG P AK+N+N AP+ ++LV+ L+
 Sbjct: 287 IGALISITGLNMGESIMVPRYGAAIADEGLLPAAIAKQONQONGAPLVAILVSGAIAIVLLL 346

-1125-

Query: 348 TFLFTQSA 355
 T F A
 Sbjct: 347 TGSFESLA 354

5

A further related DNA sequence was identified in *S.pyogenes* <SEQ ID 9079> which encodes the amino acid sequence <SEQ ID 9080>. Analysis of this protein sequence reveals the following:

Possible site: 60
 >>> Seems to have no N-terminal signal sequence

10	INTEGRAL	Likelihood = -9.92	Transmembrane	77 - 93 (72 - 100)
	INTEGRAL	Likelihood = -9.29	Transmembrane	279 - 295 (274 - 303)
	INTEGRAL	Likelihood = -9.08	Transmembrane	203 - 219 (199 - 225)
	INTEGRAL	Likelihood = -8.55	Transmembrane	174 - 190 (171 - 197)
	INTEGRAL	Likelihood = -8.33	Transmembrane	436 - 452 (432 - 455)
15	INTEGRAL	Likelihood = -7.32	Transmembrane	329 - 345 (324 - 350)
	INTEGRAL	Likelihood = -5.63	Transmembrane	402 - 418 (396 - 421)
	INTEGRAL	Likelihood = -4.88	Transmembrane	460 - 476 (456 - 479)
	INTEGRAL	Likelihood = -4.51	Transmembrane	379 - 395 (377 - 401)
	INTEGRAL	Likelihood = -2.81	Transmembrane	48 - 64 (48 - 65)
20	INTEGRAL	Likelihood = -2.23	Transmembrane	243 - 259 (243 - 262)
	INTEGRAL	Likelihood = -0.37	Transmembrane	123 - 139 (123 - 139)

----- Final Results -----
 bacterial membrane --- Certainty=0.4970(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

25

An alignment of the GAS and GBS sequences follows:

Score = 62.1 bits (148), Expect = 2e-11
 Identities = 59/250 (23%), Positives = 107/250 (42%), Gaps = 12/250 (4%)

Query: 143 WGSYLKGLLAN--YNIVLPNALNGTFNL--KNGTYIDILPV-LVMFFVTGIVLMNSKLAL 197
 WG +L L N Y +L ++L F I I+ +V++ V ++L A
 Sbjct: 95 WGYWLSAWLGNVAYAAALFSSLGYYFFKFFGNGNIIISIIGASIVIWVHFLILRGVNTAA 154

Query: 198 RFNSFLVILKFSALALFIFVGIFFIDHNNWSHFAPYGVGQITGGKTGIFAGASVMFFAFL 257
 N+ + K + +F+ + N +S +G G + + + F+
 Sbjct: 155 FINTIVTFAKLVPVLIIFLISALLAFKFNIFS-LDIWGNLHQSIQFNQVNSTMKTAVVWFI 213

Query: 258 GFESISMAVDEVKEPQKTIPKGIILSLIIVTALYIVVTTILTIGIV---HYTKLNVPDVA 314
 G E + K+ I K IL+L + +LY++++ + GI+ L P A+A
 Sbjct: 214 GIEGAVVFSGRACK-HSDIGKASILAFMTISLVLSVLSLGIMSRPELANLKTP-AMA 271

Query: 315 FALRNIRLYWAADYVSVIVAILTLITVCISMTYALARTIYSISRDLGLPKSLYTLTKKNKV 374
 + L +W A V++ I+++ ++ T A Y +++G PK + KNK
 Sbjct: 272 YVLEKAVGHWGAILVNLGVIIISVFGAILAWTLFAAELPYQAAKEGAFPK-FFAKENKNKA 330

Query: 375 PQNATLVTGL 384
 P N+ LVT L
 Sbjct: 331 PINSLLVTNL 340

50

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1013

55 A DNA sequence (GBSx1081) was identified in *S.agalactiae* <SEQ ID 3115> which encodes the amino acid sequence <SEQ ID 3116>. This protein is predicted to be unnamed protein product (argF). Analysis of this protein sequence reveals the following:

Possible site: 31
 >>> Seems to have no N-terminal signal sequence

-1126-

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.3757(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3117> which encodes the amino acid sequence <SEQ ID 3118>. Analysis of this protein sequence reveals the following:

Possible site: 31

10

>>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.48 Transmembrane 171 - 187 (171 - 188)

----- Final Results -----

15

 bacterial membrane --- Certainty=0.1192(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

20

>GP:CAB12563 GB:Z99108 similar to metabolite transporter [Bacillus subtilis]
 Identities = 190/467 (40%), Positives = 284/467 (60%), Gaps = 13/467 (2%)

25

Query: 25 TIFRKK-----KKYSNKTEMQRHFKVIDLVFLGLGSMVGTGIFTVTGIGAAKYAGPALTI 79
 ++FRKK S + R DL LG+G ++GTGIF +TG AA AGPAL I
 Sbjct: 3 SLFRKKPLETLSAQSKSKSLARTLSAFDLTLGLGICVIGTGIFVITGTVAATGAGPALII 62

30

Query: 80 SIIISAIAIGILALFYAEFASRMPSNGGAYSIVYATLGEFPAWLVGWYIIMEFLTAISSV 139
 S I++ +A + A YAEF+S +P +G YSY Y TLGE A+L+GW +++E++ A+S+V
 Sbjct: 63 SFILAGLACALAAFCYAEFSSSIPISGSVYSYVTLGELLAFLIGWDLMLLEYVIALSAV 122

35

Query: 140 AVGWGSYLKGLLANYNIVLPNALNGTFNLKNGTYIDILEPVLVMFFVTGIVLMNSKLALRF 199
 A GW SY + LLA +N+ +P AL G G ++ +++ +T IV K + RF
 Sbjct: 123 ATGWSSYFQSLLAGFNLHIPAALTGAPGSMAGAVFNLPAAVIILLITAIIVSRGVKESTRF 182

40

Query: 200 NSFLVILKFSALALFIFVGIFFDHNNWSHFAPYGVGQITGGKTGIFAGASVMFFAFLGF 259
 N+ +V++K + + LFI VGI ++ +NWS F P+G+ G+ A+ +FFA+LGF
 Sbjct: 183 NNVIVLMKIAIILLFIIVGIGYVVKPDNWSPFMPFGM-----KGVILSAATVFFAYLGF 235

45

Query: 260 ESISMAVDEVKEPQKTIPKGIILSLIIVTALYIVVTIITGIVHYTKLNVDPDAVAFALRN 319
 +++S A +EVK PQK +P GII +L + T LYI V+ +LTG++ Y KLVN D V+PAL+
 Sbjct: 236 DAVSNASEEVKNPQKNMPVGIISALAVCTVLYIAVSLVLTGMMPYAKLVNGDPVSFALKF 295

50

Query: 320 IRLYWAADYVSIVAILTLITVCISMTYALARTIYISISRDGLLPKSLYTLTKKNKVPQNAT 379
 + A +S+ AI+ + TV +++ YA R +++SRDGLLP + K P T
 Sbjct: 296 VGQDAVAGIISVGAIIGITTVMLALLYAQVRLTFAMSRDGLLPGLFAKVHPSFKTPFRNT 355

Query: 380 LVTGLLAMICAGIFPLSSLAEFVNICTLAYLIILSGAIKLRRIEGEPKANEFKTPLPVPF 439
 +TG++A AG L +LA VN+ TLA ++S A+I LR+ E KA+ F+ P VP
 Sbjct: 356 WLTGIVAAGIAGFINLGTLAHLVNMGTAAFTVISAIVLVRKKHPEIKAS-FRVPFVPV 414

Query: 440 LPMLAIIICLSFMSQYKFTWIAFAIATIIGTILIYLAYGYTHSIENK 486
 +P+++ ICL FM TW++F I +GTL+Y Y HS+ NK
 Sbjct: 415 VPIISAGICLWFMYSPLPGVTWLSFVIWIAVGTLYVFLYSRKHSLLNK 461

55 An alignment of the GAS and GBS proteins is shown below.

Identities = 312/337 (92%), Positives = 324/337 (95%)

60

Query: 1 MTQVFQGRSFLAEKDFSRREEFEYLIDFSAHLKDLKKRGVPHHYLEGKNIALLFECTSTRT 60
 MTQVFQGRSFLAEKDF+R E EYLIDFSAHLKDLKKRGVPHHYLEGKNIALLFECTSTRT
 Sbjct: 1 MTQVFQGRSFLAEKDFTRAEEYLIDFSAHLKDLKKRGVPHHYLEGKNIALLFECTSTRT 60

Query: 61 RAAFTTAAIDLGAHPEYLGANDIQLGKKESTEDTAKVLGRMFDGIEFRGFSQRMVEELAE 120
 RAAFTTAAIDLGAHPEYLGANDIQLGKKESTEDTAKVLGRMFDGIEFRGFSQRMVEELAE
 Sbjct: 61 RAAFTTAAIDLGAHPEYLGANDIQLGKKESTEDTAKVLGRMFDGIEFRGFSQRMVEELAE 120

-1127-

5 Query: 121 FSGVPVWNGLTDEWHPTQMLADYLTIKENFGKLEGITLVYCGDGRNNVANSLLVAGTLMG 180
 FSGVPVWNGLTDEWHPTQMLADY T+KENFGKLEG+TLVYCGDGRNNVANSLLV G ++G
 Sbjct: 121 FSGVPVWNGLTDEWHPTQMLADYFTVKENFGKLEGLTLVYCGDGRNNVANSLLVTGAILG 180

10 Query: 181 VNVHIFSPKELFPAAEEIVKLAEEYAKESGAHVLTNDVDEAVKGADVFTYTDVWVSMGEED 240
 VNVHIFSPKELFP EEIV LAE YAKESGA +L+T++ DEAVKGADV YTDVWVSMGEED
 Sbjct: 181 VNVHIFSPKELFP EEIV TLAEGYAKESGARILITEDADEAVKGADVLYTDVWVSMGEED 240

15 Query: 241 KFKERVELLQPYQVNMELIKKANNDNLIFLHCLPAFHDINTVYGKDVAEKFGVKEMEVD 300
 KFKERVELLQPYQVNM+L++KA ND LIFLHCLPAFHDINTVYGKDVAEKFGVKEMEVD
 Sbjct: 241 KFKERVELLQPYQVNM DLVQKAGNDKLIFLHCLPAFHDINTVYGKDVAEKFGVKEMEVD 300

Query: 301 EVFRSKYARHFDQAENRMHTIKAVMAATLGNLFIPKV 337
 EVFRSKYARHFDQAENRMHTIKAVMAATLGNLFIPKV
 Sbjct: 301 EVFRSKYARHFDQAENRMHTIKAVMAATLGNLFIPKV 337

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

20 Example 1014

A DNA sequence (GBSx1082) was identified in *S.agalactiae* <SEQ ID 3119> which encodes the amino acid sequence <SEQ ID 3120>. Analysis of this protein sequence reveals the following:

25 Possible site: 24
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0456(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

30 A related GBS nucleic acid sequence <SEQ ID 10921> which encodes amino acid sequence <SEQ ID 10922> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3121> which encodes the amino acid sequence <SEQ ID 3122>. Analysis of this protein sequence reveals the following:

35 Possible site: 61
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -5.41 Transmembrane 121 - 137 (118 - 140)

40 ----- Final Results -----
 bacterial membrane --- Certainty=0.3166(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

45 An alignment of the GAS and GBS proteins is shown below.

Identities = 65/113 (57%), Positives = 83/113 (72%)

50 Query: 31 MEEEFDDNDEQDTIYAVLYDGKQPVSTGRFLPETQTPEARLTRIATLKG YRGNGYGTKIII 90
 M ++FD NDE T+YAV+YD QPVSTG+FL ET+ EARLTRI TL Y G GYG K+
 Sbjct: 1 MADKFDANDETRTVYAVVYDNDQPVSTGQFLAETKIEARLTRIVTLADYCGCGYGA KVTE 60

Query: 91 ALENYAKENG YHYLTIHAELTAKDFYQTLGYQATGNIYMEDGEACQTLEKYLI 143
 ALE Y + G++ LTIH+ELTA+ FY+ LGYQ+ G +EDGE CQ+L K ++
 Sbjct: 61 ALETYTRREGFYQLTIHSELTAQTFYENLGYSYGP KCLEDGEYCQSLAKTIL 113

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1128-

Example 1015

A DNA sequence (GBSx1083) was identified in *S.agalactiae* <SEQ ID 3123> which encodes the amino acid sequence <SEQ ID 3124>. Analysis of this protein sequence reveals the following:

5 Possible site: 58
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2160(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 10 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3125> which encodes the amino acid sequence <SEQ ID 3126>. Analysis of this protein sequence reveals the following:

15 Possible site: 58
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2730(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 20 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 375/411 (91%), Positives = 395/411 (95%), Gaps = 1/411 (0%)

25 Query: 1 MTQTHPIHVFSEIGKLLKVMLEHPRGKEIENLMPDYLERLLFDDIPFLEDAQKEHDAFAQA 60
 MT PIHV+SEIGKLLKV+LHRPGKEIENLMPDYLERLLFDDIPFLEDAQKEHDAFAQA
 Sbjct: 1 MTAQTPIHVYSEIGKLLKVLHHRPGKEIENLMPDYLERLLFDDIPFLEDAQKEHDAFAQA 60

30 Query: 61 LRNEGVEVLYLENLAESLTNQEIREFIDEYIGEANVRGRATKKAIRELLLNKIDNKEL 120
 LR+EG+EVLYLE LAESL EIRE FIDEY+ EAN+RGRATKKAIRELL+ I+DN+EL
 Sbjct: 61 LRDEGIEVLYLETLAESLVTPEIREAFIDEYLSEANIRGRATKKAIRELLMAIEDNQEL 120

35 Query: 121 IEKTMAGIQKSELPEIPSEKGLTDLVESNYPFAIDPMPNLYFTRDPFATIGNGVSLNHM 180
 IEKTMAG+QKSELPEIP+SEKGLTDLVESNYPFAIDPMPNLYFTRDPFATIG GVS LNHM
 Sbjct: 121 IEKTMAGVQKSELPEIPASEKGLTDLVESNYPFAIDPMPNLYFTRDPFATIGTGVS LNHM 180

40 Query: 181 FSETRNRETLYGKYIFTHHPYGG-KVPMVYEREETTRIEGGDELVLSDVLA VGISQRT 239
 FSETRNRETLYGKYIFTHHP YGG KVP MVY+R ETTRIEGGDELVLSDVLA VGISQRT
 Sbjct: 181 FSETRNRETLYGKYIFTHHPYGGGKVP MVYDRNETTRIEGGDELVLSDVLA VGISQRT 240

45 Query: 240 DAASIEKLLVNIFKQNLGFKKVLAFEFANNRKFHMLDTVFTMVDYDKFTIHPEIEGDLRV 299
 DAASIEKLLVNIFKQNLGFKKVLAFEFANNRKFHMLDTVFTMVDYDKFTIHPEIEGDLRV
 Sbjct: 241 DAASIEKLLVNIFKQNLGFKKVLAFEFANNRKFHMLDTVFTMVDYDKFTIHPEIEGDLRV 300

50 Query: 300 YSVTYENQDLHIEEEKGDLADLLAKNLGVEKVELIRCGGDNLVAAGREQWNDGSNTLTIA 359
 YSVTY+N++LHI EEKGD LA+LLA NLGVEKV+LIRCGGDNLVAAGREQWNDGSNTLTIA
 Sbjct: 301 YSVTYDNEELHIVEEEKGDLAELLAA NLGVEKVDLIRCGGDNLVAAGREQWNDGSNTLTIA 360

 Query: 360 PGVVIVYNRNTITNAILES KGLKLIKINGSELVRGRGGPRC MSPFEREDL 410
 PGVV+VYNRNTITNAILES KGLKLIKI+GSELVRGRGGPRC MSPFERED+
 Sbjct: 361 PGVVVVYNRNTITNAILES KGLKLIKIHGSELVRGRGGPRC MSPFEREDI 411

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1016

A DNA sequence (GBSx1084) was identified in *S.agalactiae* <SEQ ID 3127> which encodes the amino acid sequence <SEQ ID 3128>. Analysis of this protein sequence reveals the following:

Possible site: 20

-1129-

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.3162 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 8703> which encodes amino acid sequence <SEQ ID 8704> was also identified. This protein has an RGD motif and has homology with the following sequences in the GENPEPT database.

>GP:AAG07568 GB:AE004834 hypothetical protein [Pseudomonas aeruginosa]
 Identities = 42/132 (31%), Positives = 74/132 (55%), Gaps = 3/132 (2%)

15 Query: 35 IQTYRKAYQTFKTK-KGARSSIEALLKRVNSGNEITSINPLVDIYNAASLRFGLPIGAED 93
 + + +A++ F K + S EAL KR + SI+P+VD+YNA S++F +P+G E+
 Sbjct: 63 LAAWAEAFRRFGAKPQRTPCSAEALRKRALRDGGLPSIDPVVDLYNAISVQFAIPVGGEN 122

20 Query: 94 SDTFRGDLKLTITNGGDEFYLI--GEDFNRPTLSGELAYVDDVGAVCRCFNWRDGRKRTMI 151
 + G +L + +G + F + GE + GE+ + DD+G CR +NWR G RT +
 Sbjct: 123 LAAYAGPPRLVVDGSETFDTLKNGEALDESPDPGEVWVRDDLGVTCRRWNWRQGVTRL 182

25 Query: 152 TDNTQNAFLVIE 163
 + + + ++E
 Sbjct: 183 DASARRMWFIE 194

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3129> which encodes the amino acid sequence <SEQ ID 3130>. Analysis of this protein sequence reveals the following:

Possible site: 33
 >>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0700 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 127/199 (63%), Positives = 155/199 (77%)

40 Query: 8 ELKQLLSDSHSLAKKYLQEKBFQNRVIQTYRKAYQTFKTKKGARSSIEALLKRVNSGNE 67
 ++KQLL+DSH LAK YL FS N+V+Q YRKAYQ FKTGGARSSIEALLKRV++G
 Sbjct: 36 DVKQLLADSHELAKAYLTADNFSDNQVQVYRKAYQHFKTKKGARSSIEALLKRVSNQGS 95

45 Query: 68 ITSINPLVDIYNAASLRFGLPIGAEDSDTFRGDLKLTITNGGDEFYLIIGEDFNRPTLSGE 127
 I SINPLVDIYNAASLRFGLP GAEDSD+F GDL+LTIT+GGD+FYLIG+ N PTL E
 Sbjct: 96 IPSINPLVDIYNAASLRFGLPAGAEDSDSFIGDLRLTITDGGDDFYLIGDADNNPTLPNE 155

50 Query: 128 LAYVDDVGAVCRCFNWRDGRKRTMITDNTQNAFLVIELIDNGREIIFKEALDFIATNTNRF 187
 L Y DD+GA CRC NWRDG+RTM+T++T+NAFL+IE +D + +EAL FI + +
 Sbjct: 156 LCYKDDIGAFRCCLNWRDGERIMVTEHTKNAFLIIEALDQEGQNRLQKALFKIEGSAKMY 215

55 Query: 188 LKAKTQTIILDKEHSEITL 206
 L A T +LDK++ + L
 Sbjct: 216 LHAITSVHVLDKDNPHVPL 234

SEQ ID 8704 (GBS298) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 44 (lane 2; MW 29kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 48 (lane 5; MW 54kDa).

-1130-

The GBS298-GST fusion product was purified (Figure 203, lane 9) and used to immunise mice. The resulting antiserum was used for FACS (Figure 297), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1017

A DNA sequence (GBSx1085) was identified in *S.agalactiae* <SEQ ID 3131> which encodes the amino acid sequence <SEQ ID 3132>. Analysis of this protein sequence reveals the following:

```

Possible site: 26
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3770 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1018

A DNA sequence (GBSx1086) was identified in *S.agalactiae* <SEQ ID 3133> which encodes the amino acid sequence <SEQ ID 3134>. Analysis of this protein sequence reveals the following:

```

Possible site: 20
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.4263 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:CAB95946 GB:Y17554 Crp/Fnr family protein [Bacillus
    licheniformis]
Identities = 85/214 (39%), Positives = 126/214 (58%), Gaps = 14/214 (6%)

Query: 11  RQLDDFKHFTIEQFDHIVSHIKHRTALKNHTLFFEGDYREKLFLIQSGHVKIEQSDASGS 70
           R L+D K F             I R+ K   L F E D RE+++L+ G +K+E+S+ +GS
Sbjct: 22  RDLEDKMQF-----IYWRSYHKGQILFMEDDPRERMYLLLDGFIKLEKSNEAGS 70

Query: 71  FIYTDYVRQGTVPFPGGLFLDDDDYHPSAVAITDIEYFSLPMALYEEYSLQINQMHLCR 130
           YTDYVR T+FP+GGLF D+ YH++A A+TDIE + +PM ++E+ N N + +
Sbjct: 71  MFYTDYVRPHTLFPFGGLFRDEHYHYAAEALTDIELYIIPMNIFEDLVRDNKNLLYDILN 130

Query: 131 KYSKLLRVHEIRLRNMVTSASMRVIQSLATL---LLQVPTERGHLPFPITTIEIANMSG 187
           S +L +HE RL+ + S A RV Q++ L L Q + + PIT EIA +SG
Sbjct: 131 HLSDILALHEERLKRITLSHAHDRVTAIYYLTESLGQKESNSTVINCPITAAETAKISG 190

Query: 188 TTRETIVSHVLKELRQKDIVEMKGKLLYNNKNYF 221
           T+RETVS VLK+LR + ++ K+++ N YF
Sbjct: 191 TSRETIVSAVLKKLRCEGVISQMNKQIMINRPEYF 224

```

-1131-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3135> which encodes the amino acid sequence <SEQ ID 3136>. Analysis of this protein sequence reveals the following:

Possible site: 20
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm	---	Certainty=0.4478(Affirmative)	< succ>
bacterial membrane	---	Certainty=0.0000(Not Clear)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 130/224 (58%), Positives = 180/224 (80%)

```

Query: 1  MITKEQYFYFRQLDDFKHFTIEQFDHIVSHIKHRTALKNHTLFFEGDYREKFLIQSGHV 60
      +I +E Y Y R+L+DF++F+IEQPD IV  ++ R A K+H LFFEGD R+KLFL+ SG+
Sbjct: 1  VIRREDYQYLRKLNDFRYFSIEQFDKIVGQMEFRKAKKDHLFFEGDKRDKLFLVTSGYF 60

Query: 61  KIEQSDASGSFIYTDYVRQGTVPFYGGFLDDDDYHFSVAITDIEYFSLPMALYEEYSIQ 120
      K+EQSD SG+F+YTD++R GT+FPYGGFL DD YHFS VA+TD+ YF P+ L+E+YSL+
Sbjct: 61  KVEQSDQSGTFMYTDFIRHGTFIPYGGFLTDDYHFSVAMTDVTYFYFVDFLFDYSLE 120

Query: 121 NINQMKNLCKRKYSKLLRVHEIRLRNMVTSSASMRVIQSLATLLQVPTERGHLPFPITTI 180
      N QMKHL K SKLL +HE+R+RN++TSSAS RVIQSLA LL+++ + LPF +TT
Sbjct: 121 NRLQMKHLYSKMSKLLLELHELVRNLTSSASSRVIQSLAILLVEMGKDSDTLPFQLTTT 180

Query: 181 EIANMSGTTRETVSHVLKELRQKDIVEMKGKLLYNNKNYFKKF 224
      +IA +SGTTRETVSHVL++L+++++ +KGK L Y +K+YF ++
Sbjct: 181 DIAQISGTTRETVSHVLRDLKKQELITIKGKYLTLYLDKDYFLQY 224
  
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1019

A DNA sequence (GBSx1087) was identified in *S.galactiae* <SEQ ID 3137> which encodes the amino acid sequence <SEQ ID 3138>. Analysis of this protein sequence reveals the following:

Possible site: 53
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm	---	Certainty=0.1643(Affirmative)	< succ>
bacterial membrane	---	Certainty=0.0000(Not Clear)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2161> which encodes the amino acid sequence <SEQ ID 2162>. Analysis of this protein sequence reveals the following:

Possible site: 59
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm	---	Certainty=0.1201(Affirmative)	< succ>
bacterial membrane	---	Certainty=0.0000(Not Clear)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 461/493 (93%), Positives = 478/493 (96%)

```

Query: 2  MSNWDTKFLKKGFTFDDVLLIPAESHVLPNEVDMKTKLADNLTNLNIPITTAAMDTVTDSK 61
      MSNWDTKFLKKG+TFDDVLLIPAESHVLPNEVD+KTKLADNLTNLNIPITTAAMDTVT SK
  
```

-1132-

Sbjct: 1 MSNWDTKFLKKGYTFDDVLLIPAESHVLPNEVDLKTCLADNLTLNIPITTAAMDTVTSK 60
 Query: 62 MAIAIARAGGLGIHKNMSIVDQAEVRKVKRSENGVIIDPFFLTPDNTVSEAEELMQNY 121
 MAIAIARAGGLG+IHKNMSI +QAEVRKVKRSENGVIIDPFFLTP++ VSEAEELMQ Y
 5 Sbjct: 61 MAIAIARAGGLGVHKNMSITEQAEVRKVKRSENGVIIDPFFLTPEHKVSEAEELMQRY 120
 Query: 122 RISGVPIVETLENRKLVGIITNRDMRFISDYKQLISEHMTSONLVTAPIGTDLETAERIL 181
 RISGVPIVETL NRKLVGIITNRDMRFISDY ISEHMTS++LVTA +GTDLETAERIL
 10 Sbjct: 121 RISGVPIVETLANRKLVGIITNRDMRFISDYNAPISEHMTSEHLVTAAGVTDLTAERIL 180
 Query: 182 HEHRIEKLPLVDDEGRSLGLITIKDIEKVIEFPKAAKDEFGRLLVAGAVGVTSDTFERAE 241
 HEHRIEKLPLVD+ GRSLGLITIKDIEKVIEFP AAKDEFGRLLVA AVGVTSDTFERAE
 Sbjct: 181 HEHRIEKLPLVDNSGRSLGLITIKDIEKVIEFPKAAKDEFGRLLVAAAVGVTSDTFERAE 240
 15 Query: 242 ALFEAGADAIVIDTAHGHSAGVLRKIAEIRAHFPNRTLIAIGNIATAEGARALYDAGVDVV 301
 ALFEAGADAIVIDTAHGHSAGVLRKIAEIRAHFPNRTLIAIGNIATAEGARALYDAGVDVV
 Sbjct: 241 ALFEAGADAIVIDTAHGHSAGVLRKIAEIRAHFPNRTLIAIGNIATAEGARALYDAGVDVV 300
 Query: 302 KVGIGPGSICITTRVVAGVGVQPQITAIYDAAAVAREYGKTIADGGIKYSGDIVKALAAGG 361
 KVGIGPGSICITTRVVAGVGVQPQ+TAIYDAAAVAREYGKTIADGGIKYSGDIVKALAAGG
 20 Sbjct: 301 KVGIGPGSICITTRVVAGVGVQPQVTAIYDAAAVAREYGKTIADGGIKYSGDIVKALAAGG 360
 Query: 362 NAVMLGSMFAGTDEAPGETEIFQGRKFPTYRGMGSIAMKKGSSDRYFQGSVNEANKLVP 421
 NAVMLGSMFAGTDEAPGETEI+QGRKFPTYRGMGSIAMKKGSSDRYFQGSVNEANKLVP
 25 Sbjct: 361 NAVMLGSMFAGTDEAPGETEIIYQGRKFPTYRGMGSIAMKKGSSDRYFQGSVNEANKLVP 420
 Query: 422 EGIEGRVAYKGSVADIVFQMLGGIRSGMGYVGAANIKELHDNAQFVEMSGAGLKESHPHD 481
 EGIEGRVAYKG+ +DIVFQMLGGIRSGMGYVGA +I+ELH+NAQFVEMSGAGL ESHPHD
 30 Sbjct: 421 EGIEGRVAYKGAASDIVFQMLGGIRSGMGYVAGDIQELHENAQFVEMSGAGLIESHPHD 480
 Query: 482 VQITNEAPNYSVH 494
 VQITNEAPNYSVH
 Sbjct: 481 VQITNEAPNYSVH 493
 35 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 vaccines or diagnostics.

Example 1020

A DNA sequence (GBSx1089) was identified in *Sagalactiae* <SEQ ID 3139> which encodes the amino
 acid sequence <SEQ ID 3140>. This protein is predicted to be MutR. Analysis of this protein sequence
 40 reveals the following:

Possible site: 23
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 45 bacterial cytoplasm --- Certainty=0.1841(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

50 >GP:AAD04237 GB:AF007761 MutR [Streptococcus mutans]
 Identities = 51/215 (23%), Positives = 102/215 (46%), Gaps = 9/215 (4%)
 Query: 5 GKILKELREDKGISLSSLAQAQLSKSTLSRFENGETQIGIDKFIKALQTLEVGVTTINEV 64
 G++ KELR +G+ L +A+ LS S LS+FENG+T + DK I A+Q + +T +E
 55 Sbjct: 9 GELYKELRMARGKLKLDIARD-NLSVSQLSKFENGQTMLAADKLILAIQGIH--MTFSEF 65
 Query: 65 SILDSKVKAGTSNTDLEQLTLLSEYRDNEDIMRIFSFKQQQSCDRIESNVLKILAKLFIS 124
 S ++ + ++L L++ +D + + +I + + + K++ K +
 60 Sbjct: 66 SYAFTQYQESDLFKTGKCLVELQTKDKIKGLKKILKDYPTETYNVYNRLNKLVIKAAVY 125
 Query: 125 NLGLNMRLPQDEINLVVITYLNGVTQYNDFYFKVICYFQDILPED--VILNKI----SNMT 178

-1133-

```

      +L + + +E + +YL + ++ ++ +      IL +D V L K      +
Sbjct: 126 SLDSSFETITNEEKEFLTSTYLYAIEBWTBYELYLFGNTLFLSDDDLVFLGKAFVERDKLY 185

Query: 179 KEQLPYSKSLVNLLIKQVITALEKDSVDKAI VFAD 213
      +E + K +LI ++I +E S A F +
Sbjct: 186 RELSEHKKRAELVLINLILILVEHHSFYHAQYFIE 220

```

There is also homology to SEQ ID 628.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1021

A DNA sequence (GBSx1090) was identified in *S.agalactiae* <SEQ ID 3141> which encodes the amino acid sequence <SEQ ID 3142>. Analysis of this protein sequence reveals the following:

```

Possible site: 15
>>> Seems to have a cleavable N-term signal seq.
INTEGRAL    Likelihood = -10.77    Transmembrane  269 - 285 ( 265 - 287)
INTEGRAL    Likelihood = -6.90     Transmembrane   33 - 49 ( 31 - 51)
INTEGRAL    Likelihood = -6.79     Transmembrane  182 - 198 ( 176 - 200)
INTEGRAL    Likelihood = -6.37     Transmembrane  117 - 133 ( 113 - 135)
INTEGRAL    Likelihood = -5.57     Transmembrane  240 - 256 ( 232 - 259)
INTEGRAL    Likelihood = -3.40     Transmembrane  223 - 239 ( 220 - 239)
INTEGRAL    Likelihood = -0.96     Transmembrane   56 - 72 ( 55 - 72)

----- Final Results -----
bacterial membrane --- Certainty=0.5310(Affirmative) < succ>
bacterial outside  --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3143> which encodes the amino acid sequence <SEQ ID 3144>. Analysis of this protein sequence reveals the following:

```

Possible site: 48
>>> Seems to have an uncleavable N-term signal seq
INTEGRAL    Likelihood = -10.99    Transmembrane  269 - 285 ( 264 - 286)
INTEGRAL    Likelihood = -8.76     Transmembrane  117 - 133 ( 112 - 135)
INTEGRAL    Likelihood = -7.70     Transmembrane  179 - 195 ( 174 - 200)
INTEGRAL    Likelihood = -4.83     Transmembrane   34 - 50 ( 32 - 52)
INTEGRAL    Likelihood = -4.46     Transmembrane  213 - 229 ( 211 - 230)
INTEGRAL    Likelihood = -4.14     Transmembrane  240 - 256 ( 232 - 259)
INTEGRAL    Likelihood = -0.69     Transmembrane   91 - 107 ( 91 - 108)
INTEGRAL    Likelihood = -0.32     Transmembrane    4 - 20 ( 4 - 20)

----- Final Results -----
bacterial membrane --- Certainty=0.5394(Affirmative) < succ>
bacterial outside  --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related sequence was also identified in GAS <SEQ ID 9181> which encodes the amino acid sequence <SEQ ID 9182>. Analysis of this protein sequence reveals the following:

```

Possible site: 38
>>> Seems to have an uncleavable N-term signal seq
INTEGRAL    Likelihood = -10.99    Transmembrane  259 - 275 ( 254 - 276)
INTEGRAL    Likelihood = -8.76     Transmembrane  107 - 123 ( 102 - 125)
INTEGRAL    Likelihood = -7.70     Transmembrane  169 - 185 ( 164 - 190)
INTEGRAL    Likelihood = -4.83     Transmembrane   24 - 40 ( 22 - 42)
INTEGRAL    Likelihood = -4.46     Transmembrane  203 - 219 ( 201 - 220)
INTEGRAL    Likelihood = -4.14     Transmembrane  230 - 246 ( 222 - 249)
INTEGRAL    Likelihood = -0.69     Transmembrane   81 - 97 ( 81 - 98)

```

-1134-

----- Final Results -----

bacterial membrane --- Certainty=0.539(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

5

An alignment of the GAS and GBS proteins is shown below.

Identities = 200/287 (69%), Positives = 244/287 (84%)

10 Query: 1 MEGLLIALIPMFAWGSIGFVSNKIGGRPNQQTFGMTLGALLFAIIVWLFKQPEMTASLWI 60
 +EG+ ALIPMF WGSIGFVSNKIGG+P+QQT GMT GALLF++ VWL +PEMT LW+
 Sbjct: 1 LEGIFYALIPMFTWGSIGFVSNKIGGKPSQQTGMTFGALLFSLAVWLIVRPEMTLQLWL 60

15 Query: 61 FGILGGILWSVGQNGQFQAMKYMGVSVANPLSSGAQLVGGSLVGALVFEHWTKPIQFILG 120
 FGILGG +WS+GQ GQF AM+YMGVSVANPLSSG+QLV GSL+G LVFEHWT+P+QF++G
 Sbjct: 61 FGILGGFIWSIGQTGQFHAMQYMGVSVANPLSSGSQLVLGSLIGVLVFEHWTQPMQFVVG 120

20 Query: 121 LTALTLLVIGFYFSSKRDVSEQALATHQEFSSKGFATIAYSTVGYISYAVLFNNIMKFDAM 180
 AL LL++GFYFSSK+D + + FSKGF + YST+GY+ YAVLFNNIMKF+ +
 Sbjct: 121 SLALLLLVIGFYFSSKQDDANAQVNHNLNFSKGFALTYSTIGYVMYAVLFNNIMKFEVL 180

25 Query: 181 AVILPMAVGMCLGAICFMKFRVNFEAVVKNMITGLMWGVGNVFMLLAAKAGLAIAFSF 240
 +VILPMAVGM LGAI FM F+++ + V+KN + GL+WG+GN+FMLLAA+KAGLAIAFSF
 Sbjct: 181 SVILPMAVGMVLGAITFMSPKISIDQYVIKNSVVGLLWIGIGNIFMLLAAKAGLAIAFSF 240

30 Query: 241 SQLGVIISIIGILFLGETKTKKEQKVVVMGILCFVMGAILLGIVKS 287
 SQLG IISI+GGILFLGETKTKKE +WVV GI+CF++GAILLG+VKS
 Sbjct: 241 SQLGAIISIVGGILFLGETKTKKEMRVVVTGIICFIVGAILLGIVKS 287

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1022

A DNA sequence (GBSx1092) was identified in *S.galactiae* <SEQ ID 3145> which encodes the amino acid sequence <SEQ ID 3146>. This protein is predicted to be recF protein (recF). Analysis of this protein sequence reveals the following:

35 Possible site: 56
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

40 bacterial cytoplasm --- Certainty=0.2653(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3147> which encodes the amino acid sequence <SEQ ID 3148>. Analysis of this protein sequence reveals the following:

45 Possible site: 56
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

50 bacterial cytoplasm --- Certainty=0.1677(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 248/364 (68%), Positives = 300/364 (82%), Gaps = 1/364 (0%)

55

Query: 1 MWIKNISLKHRYNYBEAQVDFSPNLNIFIGRNAQGKTNFLEAIYFLALTRSHRTRSDKEL 60
 MWIK + LKHRYNY+ FS LN+FIG NAQGKTNFLEAIYFL+LTRSHRTR+DKEL
 Sbjct: 1 MWIKELELKHRYNYDHLASFSSGLNVFIGNNAQGKTNFLEAIYFLSLTRSHRTRADKEL 60

-1135-

5 Query: 61 VHFKHHDVQITGEVIRKSGHNLNDIQLSEKGRITKVNHLKQAKLSDYIGAMTVVLFAPED 120
 +HF H V +TG++ R SG ++L+I LS+KGR+TK+N LKQAKLSDYIG M VVLFAPED
 Sbjct: 61 IHFDHSTVSLTGKIQRISGTVLEINLSDKGRVTKINALKQAKLSDYIGTMMVVLFAPED 120

10 Query: 121 LQLVKGAPSLRRKFLDIDIGQIKPTYLAELSNVNHVLKQRNTYLKTTNNVDKTFLLTVLDE 180
 LQLVKGAPSLRRKF+DID+GQIKP YL+ELS+YNHVLKQRN+YLK+ +D FL VLDE
 Sbjct: 121 LQLVKGAPSLRRKFIDIDLQGIKPVYLSLSHYNHVLKQRNSYLSAQQIDAAPLAVLDE 180

15 Query: 241 QLSKSHSRDIFKNTSIGPHRDDITFFINDINATFASQGGQSLILSLKLAELIKTVT 300
 QL K+H +D F+KNTS+GPHRD++ F+IN +NA FASQGG RSLILSLK+AE+ L+K +T
 Sbjct: 241 QLEKNHQKDFFRKNTSVGPHRDELAFYINGMNNANFASQGGHRSILSLKMAEVSLMKALT 300

20 Query: 301 NDYPILLDDVMSELNHRQLKLEGG-IKENVQTFITTTSLHLSALPDQLKIFNVSDGT 359
 D PILLDDVMSELN RQ KLE IKENVQTFITTTSL+HLS LP+ ++IF+V+ GT
 Sbjct: 301 GDNPILLDDVMSELNTRQTKLETVIKENVQTFITTTSLDHLSQLPEGIRIFHVTKGT 360

25 Query: 360 ISIN 363
 + I+
 Sbjct: 361 VQID 364

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1023

30 A DNA sequence (GBSx1093) was identified in *S.galactiae* <SEQ ID 3149> which encodes the amino acid sequence <SEQ ID 3150>. Analysis of this protein sequence reveals the following:

Possible site: 26
 >>> Seems to have no N-terminal signal sequence

35 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1807(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

40 >GP:CAA61548 GB:X89367 orf121 [Lactococcus lactis]
 Identities = 56/116 (48%), Positives = 74/116 (63%), Gaps = 3/116 (2%)

Query: 3 YKLFDEYITLQSLLEIGIIQSGGAIKKFLADNR--VLFNGDLENRRGKKLRGLDITIP 60
 Y LF+EYITL LLKE+G+I +GG K FLA+N + +NG+ ENRRGKKLR GD++ P
 45 Sbjct: 4 YILFEEYITLGQLLKEGLISTGGQPKIFLAENEGNIFYNGEENRRGKKLRDGLLEFP 63

Query: 61 DQNIETIIIRKPSDQETIERNIEIAEKQRVSAIVKEMNKNTNKGSKTSKKPVRFP 116
 ++++ + I+E E AE+ RV AIVK+MN NK K P RFEP
 50 Sbjct: 64 TFDLKVTFEQADADAIKEHEAEAEARVKAIVKMNAE-NKTTKPAKKAPPRFP 118

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3151> which encodes the amino acid sequence <SEQ ID 3152>. Analysis of this protein sequence reveals the following:

Possible site: 34
 >>> Seems to have no N-terminal signal sequence

55 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0493(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

60

-1136-

An alignment of the GAS and GBS proteins is shown below.

Identities = 74/136 (54%), Positives = 94/136 (68%), Gaps = 20/136 (14%)

```

5  Query: 1  MDYKLFDEYITLQSLLEIGIIQSGGAIKKFLADNRVLFNGDLENRRGKKLRGLGDIITIP 60
      M YKLF E+ITLQ+LLKE+GIIQSGGAIK FLA+ VLFNG+ E RRGKK+R+GD I++P
Sbjct: 9  MIYKLFTEFITLQALLKELGIIQSGGAIKGFLAETTVLFNGEDEKRRGKKIRVGDKISLP 68

Query: 61  DONIEIIRKPSDQEIERNIEIAEKQVSAIVKEMNKNTNKGKSK-----TSKK---- 110
      DQ++ I I +PS +E E+ E+AEK RV+A+VK+MN+ K SK T+KK
10 Sbjct: 69  DQDLIITIVEPSQEEKEQFAEEMAETKTRVAALVKQMNQANKKTSSKHNNRQSTTKKSLRA 128

Query: 111 -----PVRFPFG 116
      PVRFPFG
15 Sbjct: 129 TKKTKGKFTAPVRFPG 144

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1024

A DNA sequence (GBSx1094) was identified in *Sagalactiae* <SEQ ID 3153> which encodes the amino acid sequence <SEQ ID 3154>. Analysis of this protein sequence reveals the following:

```

Possible site: 47
>>> Seems to have no N-terminal signal sequence
      INTEGRAL      Likelihood = -1.86      Transmembrane 269 - 285 ( 267 - 285)

25 ----- Final Results -----
      bacterial membrane --- Certainty=0.1744(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3155> which encodes the amino acid sequence <SEQ ID 3156>. Analysis of this protein sequence reveals the following:

```

Possible site: 48
>>> Seems to have no N-terminal signal sequence

35 ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3008(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

Identities = 227/413 (54%), Positives = 309/413 (73%)

```

Query: 1  MKIVEGVSLHLIKNQQFKTNHLTFRFSGDFNNKTVARSLVAQMLVTANAKYPKVQEFRE 60
      MKIV+GV LHLIK +QFKTNH+TFRFSGD N KTVA++ LVAQML TAN YP V++PRE
45 Sbjct: 1  MKIVQGVQLHLIKTKQFKTNHITFRFSGDLNQKTVAKKVLVAQMLATANECYPTVRQFRE 60

Query: 61  KLASLYGASLSTKISTKGLVHIVDIDIVFKNTFTLEQENIVEQIITFLEDMLFSPLISL 120
      KLA LYGASLST + TKGLVHIVDIDI F+++ + E I++++I FL+D+LFSPL+S+
50 Sbjct: 61  KLARLYGASLSTNVLTGGLVHIVDIDITFIQDRYACNGEKILDEMIQFLKDILFSPLLSI 120

Query: 121  EQYQTSIFDTEKKNLIQYLEADIEDNFYSSDLALKSLFYNNKTLRLPKYGTASLVESENS 180
      QYQ +F+TEK NLI Y+E+D ED+FY S L +K LFY NK L++ +YG+ L+ E +
Sbjct: 121  AQYQPKVFETKKNLINYIESDREDSFYYSLLKVKELFYCNKQLQMSYEGSPELIAKETA 180

55 Query: 181  FTAYQEFQKMLKEDQLDIFVVGDFDDYRMIQAFNRMAFEPRIHKVLAFDYDTQTYENITRSQ 240
      +T+YQEF KML EDQ+DIF++GDFDDYR++Q ++ + R+K L F + Q NI +
Sbjct: 181  YTSYQEFHKMLNEDQIDIFILGDFDDYRVVQLIHQFPLDNRNKNLNFHQLQNSVNIKES 240

Query: 241  VEDKDVNQSIMQLAYHLPTIKDEDFALIVFNGLFGAFAHSLFTIREKQGLAYTIGS 300

```

-1137-

+E + V+QSI+QLAYH P + DY+AL++ NGL G+FAHS LF +IRE++GLAY+IG
 Sbjct: 241 IEKRAVHQSIQLAYHFPSPVFGQRDYYALVLLNGLLGSFAHSRLFIKIRREEGLAYSIGC 300
 Query: 301 QFDSFTGLFTIYAGIDKENRERFLKLINKQFNNIKMGRFSSTLLKQTKDILKMNYVLASD 360
 +FDS+TGLF IY GID ++R + L+LI ++ N IKMGRFS L+K+T+ +L N +L+ D
 Sbjct: 301 RFDSTYGLFEIYTGIDSQHRITKTLQLIIQELNAIKMGRFSEQLIKKTRSMMLNALLSED 360
 Query: 361 NPKVIVDHIYHEHYLDQFHTSALFIDKVDVTKSDIVSVATKLLQAFYFLEG 413
 K I++ IY Y+D ++ +I V++V K+DI+ VA LKLQ YFLEG
 Sbjct: 361 YNKNIIERYRSSYIDSSYSIKNWIKGVNEVKNADIIVANLLKLQTVYFLEG 413

SEQ ID 3154 (GBS400) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 76 (lane 2; MW 49.2kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 84 (lane 3; MW 74kDa) and in Figure 177 (lane 6; MW 74kDa).

GBS400-GST was purified as shown in Figure 217, lane 10.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1025

A DNA sequence (GBSx1095) was identified in *S.agalactiae* <SEQ ID 3157> which encodes the amino acid sequence <SEQ ID 3158>. Analysis of this protein sequence reveals the following:

Possible site: 50
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3473(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3159> which encodes the amino acid sequence <SEQ ID 3160>. Analysis of this protein sequence reveals the following:

Possible site: 45
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4298(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 207/424 (48%), Positives = 276/424 (64%), Gaps = 3/424 (0%)
 Query: 5 KITYQNLQEEVYKLTLESGLNVYLIPKPSFKETVGVLTANFGSLHTKYTRNGCVEHYVPAG 64
 KI Y N+ E++Y + LE+GL VY I K F E +LT PGS L K T + PAG
 Sbjct: 6 KINYENIDEDLYYVKLENGLTVYFIKKIGFLEKTAMLTVGFGSLDNKLTVDDES RDAPAG 65
 Query: 65 IAHFLEHKLFE LDKGQDAATQFTKYGAESNAFTTFDKTSFYFSTISHITNCLDILLDFVL 124
 IAHFLEHKLFE + G D + +FT+ GAE+NAFTTF++TSF+FST S L++L FVL
 Sbjct: 66 IAHFLEHKLFEDES GDISLKF TQLGAETNAFTTFNQTSFFFSTASKFQENLELLQYFVL 125
 Query: 125 TTNFTEESITKEKDIIKQEIEMYQDDPEYRLYQGVLSNLYPN SPLAFDIAGDYQSISQIT 184
 + N T+ES+++EK II QEI+MYQDD +YR Y G+L NL+P + LA DIAG SI +IT
 Sbjct: 126 SANITDESVSREKKIIGQEIDMYQDDADYRAYSGILQNLFPKTS LANDIAGSKASIQKIT 185
 Query: 185 LTDLQENHKDFYQLSNMNLVLVGQFSPQEIITYLQKNSHFTSY--SQNIDRDSISLEPVI 242
 L+ +H FYQ +NM+L +VG E +Q+ SY + + D + PVI

-1138-

Sbjct: 186 KILLETHHTYFYQPTNMSLFIVGDIDIDETFLAIQRFQTTLSYPDRKRVTVDP LHYYPVI 245

Query: 243 KNNSCHMTVT KPKLAIGYRKSNNMIHGSYLKEKIGLQLFFAMLLGWTSTINQDWYESGQI 302
K++S M VT KL +G+R + S L +I L+LF +ML+GWTS I YE G+I

5 Sbjct: 246 KSSSVDMDVITAKLVVGFGRGYLTLTQHSLLT YRIALKLFLSMLIGWTSKIYHTLYEDGKI 305

Query: 303 DDSFDIEIEVHPDFECVIIISLDITEPIAFSTQLRLLLNALQSSDLTESH LKNVKRELYG 362
DDSF+++E+H +F+ V+ISLDT EPIA S +R L S + T HL +K+E+YG

10 Sbjct: 306 DDSFDVDVEIHNFQFVLISLDTPEPIAMSNYIRQKLATIKISKEFTNEHLNLLKKEMYG 365

Query: 363 DFLRSLDSIENLAMQFVTYLYDG-KTMYLDLPSIVEELDLEDVITIGKDFLDNADTSDFV 421
DF++SLDSIE+L QF YL D K Y D+P I+E L L+DV+TIGK F + AD SDF

Sbjct: 366 DFIQSLDSIEHLTHQFSLYLSDSKETYFDIPKIIERLTLKDVVTIGKAFFEKADASDFT 425

15 Query: 422 IFPK 425
+FPK
Sbjct: 426 VFPK 429

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
20 vaccines or diagnostics.

Example 1026

A DNA sequence (GBSx1096) was identified in *S.galactiae* <SEQ ID 3161> which encodes the amino acid sequence <SEQ ID 3162>. This protein is predicted to be phosphotidylglycerophosphate synthase (pgsA). Analysis of this protein sequence reveals the following:

25 Possible site: 55
>>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -8.17	Transmembrane	17 - 33 (14 - 39)
INTEGRAL	Likelihood = -3.77	Transmembrane	92 - 108 (88 - 108)
INTEGRAL	Likelihood = -2.87	Transmembrane	144 - 160 (142 - 162)
INTEGRAL	Likelihood = -1.65	Transmembrane	42 - 58 (42 - 59)

30

----- Final Results -----

bacterial membrane	---	Certainty=0.4270(Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

35

A related GBS nucleic acid sequence <SEQ ID 10293> which encodes amino acid sequence <SEQ ID 10294> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3163> which encodes the amino acid
40 sequence <SEQ ID 3164>. Analysis of this protein sequence reveals the following:

Possible site: 48
>>> Seems to have a cleavable N-term signal seq.

INTEGRAL	Likelihood = -6.64	Transmembrane	76 - 92 (72 - 102)
INTEGRAL	Likelihood = -5.36	Transmembrane	136 - 152 (131 - 164)
INTEGRAL	Likelihood = -2.34	Transmembrane	98 - 114 (97 - 114)

45

----- Final Results -----

bacterial membrane	---	Certainty=0.3654(Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000(Not Clear)	< succ>

50

An alignment of the GAS and GBS proteins is shown below.

Identities = 145/180 (80%), Positives = 160/180 (88%)

55 Query: 8 MMKKENIPNLLTVVRILMIPLFVILTSVTSTTWHIVAAIVFAIASLTDYLDGYLARKWQ 67
M+KKENIPNLLT+VRI MIP F+ +TS + WHI AA++FAIAS TDYLDGYLARKW
Sbjct: 1 MIKKENIPNLLTVRIAMIPFLLFITSSSNKVGWHIFAIVFAIASFTDYLDGYLARKWH 60

-1139-

Query: 68 VVTNFGKFADPLADKMLVMSAFIMLVGLDLAPAWVSAIIICRELAVTGLRLLLVETGGTV 127
 V +NFGKFADPLADKMLVMSAFIMLVGL L PAWVSA+IIICRELAVTGLRLLLVETGG V
 Sbjct: 61 VASNFGKFADPLADKMLVMSAFIMLVGLVLPAPWVSAVIIICRELAVTGLRLLLVETGGKV 120

5 Query: 128 LAAAMPGKIKTATQMFAVIFLLVHWMTLGNIMLYIALFFFTLYSGYDYFKGAGFLFKDTFK 187
 LAAAMPGKIKTATQM ++I LL HW+ LGN++LYIALFFT+YSGYDYFKGA FLFKDTFK
 Sbjct: 121 LAAAMPGKIKTATQMLSIIILLCHWIFLGNVLLYIALFFFTIYSGYDYFKGASFLFKDTFK 180

A related GBS gene <SEQ ID 8705> and protein <SEQ ID 8706> were also identified. Analysis of this
 10 protein sequence reveals the following:

Lipop Possible site: -1 Crend: 4
 SRCFLG: 0
 McG: Length of UR: 9
 Peak Value of UR: 3.03
 15 Net Charge of CR: 1
 McG: Discrim Score: 6.36
 GvH: Signal Score (-7.5): -0.400001
 Possible site: 48
 >>> Seems to have a cleavable N-term signal seq.
 20 Amino Acid Composition: calculated from 49
 ALOM program count: 2 value: -3.77 threshold: 0.0
 INTEGRAL Likelihood = -3.77 Transmembrane 85 - 101 (81 - 101)
 INTEGRAL Likelihood = -2.87 Transmembrane 137 - 153 (135 - 155)
 PERIPHERAL Likelihood = 1.27 109
 25 modified ALOM score: 1.25
 icml HYPID: 7 CFP: 0.251
 *** Reasoning Step: 3
 30 ----- Final Results -----
 bacterial membrane --- Certainty=0.2508(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

35 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 vaccines or diagnostics.

Example 1027

A DNA sequence (GBSx1097) was identified in *S.agalactiae* <SEQ ID 3165> which encodes the amino
 acid sequence <SEQ ID 3166>. This protein is predicted to be ABC transporter ATP-binding protein
 40 (potA). Analysis of this protein sequence reveals the following:

Possible site: 32
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 45 bacterial cytoplasm --- Certainty=0.1805(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

50 >GP:AAC61484 GB:AF082738 ABC transporter ATP-binding protein
 [Streptococcus pyogenes]
 Identities = 201/279 (72%), Positives = 231/279 (82%)
 55 Query: 1 MTNIITVNNLFFKYDSNQTHYQLENVSFHVKGQEWLSIIHNGSGKSTTVRLIDGLLEAE 60
 M+ II + + F Y +Q L+ VSFHVKGQEWLSIIHNGSGKSTT+RLIDGLLE E
 Sbjct: 18 MSATIELKRVTFFNYHKDQEKPTLDGVSFHVKGQEWLSIIHNGSGKSTTIRLIDGLLEFE 77
 Query: 61 SGQIIIDGQELTEDNVWELRHKIGMVFQNPDPNQFVGATVEDDVAFLGLENKGIPLKDMKER 120
 SG II+DG LT NVWE+RHKIGMVFQNPDPNQFVGATVEDDVAFLGLENKGI +D+KER

-1140-

Sbjct: 78 SGSIIVDGDLLTITNVWEIRHKIGMVFNQPNQFVGATVEDDVAFGLENKGIAHEDIKER 137

Query: 121 VDQALDLVGMSEFKMREPARLSSGGQKQORVAIAGAVAMRPQVIILDEATSMLDPEGRLELI 180
V+ AL+LVGM FK +EPARLSSGGQKQORVAIAGAVAM+P++IILDEATSMLDP+GRLELI

5 Sbjct: 138 VNHALELVGMQNFKEKEPARLSSGGQKQORVAIAGAVAMKPKIILDEATSMLDPKGRLELI 197

Query: 181 RTIRAIRQKYNLTVISITHDLDEVALSDRVIVMKNKGVESTSTPKALFGRGNRLISLGLD 240
+TI+ IR Y LTVISITHDLDEVALSDRV+VMK+G+VESTSTP+ LF RG+ L+ LGLD

10 Sbjct: 198 KTIKNIRDDYQLTVISITHDLDEVALSDRVLMKDGQVESTSTPEQLFARGDELLQLGLD 257

Query: 241 VPFTSRLMAELAANGLDIGTEYLTEKELEQLWELNLKM 279
+PFT+ ++ L G I YLKELE QL +L KM

Sbjct: 258 IPFTTSVVQMLQEEGYPIDYGYLTEKELENQLCQLISKM 296

- 15 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3167> which encodes the amino acid sequence <SEQ ID 3168>. Analysis of this protein sequence reveals the following:

Possible site: 19
>>> Seems to have no N-terminal signal sequence

20 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.2235(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

25 RGD motif: 247-249

An alignment of the GAS and GBS proteins is shown below.

Identities = 200/279 (71%), Positives = 231/279 (82%)

30 Query: 1 MTNIIITVNNLFFKYDSNQTHYQLENVSFHVKGGEWLSIIGHNGSGKSTTVRLIDGLLEAE 60
M+ II + + F Y +Q L+ VSFHVKGGEWLSIIGHNGSGKSTT+RLIDGLLE E
Sbjct: 18 MSATIELKKVTFNYHKDQEKPTLDGVSFHVKGGEWLSIIGHNGSGKSTTIRLIDGLLEPE 77

35 Query: 61 SGQIIIDGQELTEDNVWELRHKIGMVFNQPNQFVGATVEDDVAFGLENKGIPKDKMER 120
SG II+DG LT NVWE+RHKIGMVFNQPNQFVGATVEDDVAFGLENKGI +D+KER
Sbjct: 78 SGSIIVDGDLLTITNVWEIRHKIGMVFNQPNQFVGATVEDDVAFGLENKGIAHEDIKER 137

40 Query: 121 VDQALDLVGMSEFKMREPARLSSGGQKQORVAIAGAVAMRPQVIILDEATSMLDPEGRLELI 180
V+ AL+LVGM FK +EPARLSSGGQKQORVAIAGAVAM+P++IILDEATSMLDP+GRLELI
Sbjct: 138 VNHALELVGMQNFKEKEPARLSSGGQKQORVAIAGAVAMKPKIILDEATSMLDPKGRLELI 197

45 Query: 181 RTIRAIRQKYNLTVISITHDLDEVALSDRVIVMKNKGVESTSTPKALFGRGNRLISLGLD 240
+TI+ IR Y LTVISITHDLDEVALSDRV+VMK+G+VESTSTP+ LF RG+ L+ LGLD
Sbjct: 198 KTIKNIRDDYQLTVISITHDLDEVALSDRVLMKDGQVESTSTPEQLFARGDELLQLGLD 257

Query: 241 VPFTSRLMAELAANGLDIGTEYLTEKELEQLWELNLKM 279
+PFT+ ++ L G + YLKELE QL +L KM
Sbjct: 258 IPFTTSVVQMLQEEGYFVDYGYLTEKELENQLCQLISKM 296

- 50 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1028

A DNA sequence (GBSx1098) was identified in *S.agalactiae* <SEQ ID 3169> which encodes the amino acid sequence <SEQ ID 3170>. Analysis of this protein sequence reveals the following:

55 Possible site: 49
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -0.27 Transmembrane 154 - 170 (154 - 170)

60 ----- Final Results -----
bacterial membrane --- Certainty=0.1107(Affirmative) < succ>

-1141-

bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

5 >GP:CAB11922 GB:Z99104 similar to ABC transporter (ATP-binding
 protein) [Bacillus subtilis]
 Identities = 141/242 (58%), Positives = 188/242 (77%), Gaps = 1/242 (0%)

10 Query: 16 TPFEGRALFDVNLKIEDASYTAFIGHTGSGKSTIMQLLNGLHIPTKGEVIVDDFSIKAGD 75
 TPFE AL+D+N I++ SY A IGHGTSGKST++Q LNGL PTKG++ + I+AG
 Sbjct: 3 TPFERLALYDINASIKEGSYVAVIGHTGSGKSTLLQHLNGLLKPTKGQISLGSTVIQAGK 62

15 Query: 76 KNKEIKFIRQKVGLVVFQFPESQLFEETVLKDVAFGPQNFQISQIEAERLAEKRLRVGIS 135
 KNK++K +R+KVG+VFQFPE QLFEEETVLKD++FGP NFG+ + +AE+ A E L+LVG+S
 Sbjct: 63 KNKDLKKLRKKVGIVVFQFPEHQLFEETVLKDISFGPMNFGVKKEDAEQKAREMLQLVGLS 122

20 Query: 136 EDLFDKNPFELSGGQMRRAIAGILAMEPKVLVLDEPTAGLDPKGRKELMTLFKNLHKKG 195
 E+L D++PFELSGGQMRRAIAG+LAM+P+VLVLDEPTAGLDP+GRKE+M +F LH++G
 Sbjct: 123 EELLDRSPFELSGGQMRRAIAGVLAAMDPEVLVLDEPTAGLDPGRKEIMDMFYELHQRG 182

25 Query: 196 -MTIVLVTHLMDDVADYADYVYVLEAGKVTLSGQPKQIFQEVELLESKQLGVPKITKFAQ 254
 +T +LVTH M+D A YAD + V+ G + SG P+ +F + E + L +P+ KF +
 Sbjct: 183 NLTTILVTHSMEDAAAYADEMIVMHKGTIQASGSPRDLFLKGEEMAGWGLDLPETIKFQR 242

Query: 255 RL 256
 L
 Sbjct: 243 HL 244

30 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3171> which encodes the amino acid
 sequence <SEQ ID 3172>. Analysis of this protein sequence reveals the following:

Possible site: 40

35 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.27 Transmembrane 154 - 170 (154 - 170)

----- Final Results -----
 bacterial membrane --- Certainty=0.1107(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

45 >GP:CAB11922 GB:Z99104 similar to ABC transporter (ATP-binding
 protein) [Bacillus subtilis]
 Identities = 146/259 (56%), Positives = 187/259 (71%), Gaps = 2/259 (0%)

50 Query: 16 TPFEGRALFNINLDILDGYSYTAFIGHTGSGKSTIMQLLNGLHVPTTGIVSVDKQDITNHS 75
 TPFE AL++IN I +GSY A IGHGTSGKST++Q LNGL PT G +S+ I
 Sbjct: 3 TPFERLALYDINASIKEGSYVAVIGHTGSGKSTLLQHLNGLLKPTKGQISLGSTVIQAGK 62

55 Query: 76 KNKEIKSIRKHVGLVVFQFPESQLFEETVLKDVAFGPQNFQVSPPEAEALAREKLALVVIS 135
 KNK++K +RK VG+VFQFPE QLFEEETVLKD++FGP NFGV E+AE ARE L LVG+S
 Sbjct: 63 KNKDLKKLRKKVGIVVFQFPEHQLFEETVLKDISFGPMNFGVKKEDAEQKAREMLQLVGLS 122

60 Query: 136 ENLFEKNPFELSGGQMRRAIAGILAMQPKVLVLDEPTAGLDPKGRKELMTIFKKLHQSG 195
 E L +++PFELSGGQMRRAIAG+LAM P+VLVLDEPTAGLDP+GRKE+M +F +LHQ G
 Sbjct: 123 EELLDRSPFELSGGQMRRAIAGVLAAMDPEVLVLDEPTAGLDPGRKEIMDMFYELHQRG 182

Query: 196 -MTIVLVTHLMDDVANYADFVYVLDKSKIILSGKPKTIFQQVSLLEKKQLGVPKVTKLAQ 254
 +T +LVTH M+D A YAD + V+ KG I SG P+ +F + + L +P+ K +
 Sbjct: 183 NLTTILVTHSMEDAAAYADEMIVMHKGTIQASGSPRDLFLKGEEMAGWGLDLPETIKFQR 242

Query: 255 RL-VDRGIPISSLPITLEE 272
 L G+ + +T+E+
 Sbjct: 243 HLEAALGVRFNEPMLTIED 261

-1142-

An alignment of the GAS and GBS proteins is shown below.

Identities = 218/280 (77%), Positives = 241/280 (85%)

```

5  Query: 1  MGIEFKNVSYTYQAGTPFEGRALFDVNLKIEDASYTAFIGHTGSGKSTIMQLLNGLHIPT 60
      M I  +NVSYTYQAGTPFEGRALF++NL I D SYTAFIGHTGSGKSTIMQLLNGLH+PT
Sbjct: 1  MSINLQNVSYTYQAGTPFEGRALFNINLDILDGSYTAFIGHTGSGKSTIMQLLNGLHVPT 60

10 Query: 61  KGEVIVDDFSIKAGDKNKEIKFIRQKVGLVFQFPESQLFEETVLKDVAFGPQNFGISQIE 120
      G V VD  I  KNKEIK IR+ VGLVFQFPESQLFEETVLKDVAFGPQNFG+S E
Sbjct: 61  TGIVSVDKQDITNHSKNKEIKSIRKHVGLVFQFPESQLFEETVLKDVAFGPQNFGVSPEE 120

15 Query: 121 AERLAEEKLRLVGISEDLFDKNPFELSGGQMRRAIAGILAMEPKVLVLDEPTAGLDPKG 180
      AE LA EKL LVGISE+LF+KNPFELSGGQMRRAIAGILAM+PKVLVLDEPTAGLDPKG
Sbjct: 121 AEALAREKLALVGISENLFEKNPFELSGGQMRRAIAGILAMQPKVLVLDEPTAGLDPKG 180

20 Query: 181 RKELMTLTFKNLHKKGMTIVLVTHLMDDVDYADYVYVLEAGKVTLSGQPKQIFQEVELLE 240
      RKELMT+FK LH+ GMTIVLVTHLMDDVA+YAD+VYVL+ GK+ LSG+PK IFQ+V LLE
Sbjct: 181 RKELMTIFKKLHQSGMTIVLVTHLMDDVANYADFVYVLDKGKIIISGKPKTIFQQVSILE 240

25 Query: 241 SKQLGVPKITTKFAQRLSHKGLNLPISLPITINEFVEAIKHG 280
      KQLGVPK+TK AQLR  +G+ + SLPIT+ E E +KHG
Sbjct: 241 KKQLGVPKVTKLAQRLVDRGIPISSLPITLEELREVLKHG 280

```

25 SEQ ID 3170 (GBS401) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 76 (lane 3; MW 34.4kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 84 (lane 4; MW 59kDa).

GBS401-GST was purified as shown in Figure 218, lane 2.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1029

A DNA sequence (GBSx1099) was identified in *S.galactiae* <SEQ ID 3173> which encodes the amino acid sequence <SEQ ID 3174>. Analysis of this protein sequence reveals the following:

```

35 Possible site: 43
    >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -10.46    Transmembrane    47 - 63 ( 25 - 69)
      INTEGRAL    Likelihood = -8.81     Transmembrane    252 - 268 ( 249 - 269)
      INTEGRAL    Likelihood = -7.91     Transmembrane    116 - 132 ( 110 - 141)
40  INTEGRAL    Likelihood = -4.25     Transmembrane    29 - 45 ( 25 - 46)
      INTEGRAL    Likelihood = -2.55     Transmembrane    77 - 93 ( 77 - 95)
      INTEGRAL    Likelihood = -0.43     Transmembrane    199 - 215 ( 199 - 215)

----- Final Results -----
45      bacterial membrane --- Certainty=0.5182(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 8707> which encodes amino acid sequence <SEQ ID 8708> was also identified. Analysis of this protein sequence reveals the following:

```

50 Lipop: Possible site: -1    Crend: 7
    SRCFLG: 0
    McG: Length of UR: 8
      Peak Value of UR: 0.65
      Net Charge of CR: 1
55 McG: Discrim Score: -10.55
    GvH: Signal Score (-7.5): 1.45

```


-1143-

Possible site: 37
 >>> Seems to have no N-terminal signal sequence
 Amino Acid Composition: calculated from 1
 ALOM program count: 6 value: -10.46 threshold: 0.0

5	INTEGRAL	Likelihood = -10.46	Transmembrane	41 - 57 (19 - 63)
	INTEGRAL	Likelihood = -8.81	Transmembrane	246 - 262 (243 - 263)
	INTEGRAL	Likelihood = -7.91	Transmembrane	110 - 126 (104 - 135)
	INTEGRAL	Likelihood = -4.25	Transmembrane	23 - 39 (19 - 40)
	INTEGRAL	Likelihood = -2.55	Transmembrane	71 - 87 (71 - 89)
10	INTEGRAL	Likelihood = -0.43	Transmembrane	193 - 209 (193 - 209)
	PERIPHERAL	Likelihood = 0.79		90

modified ALOM score: 2.59
 icml HYPID: 7 CFP: 0.518

15 *** Reasoning Step: 3

----- Final Results -----

	bacterial membrane --- Certainty=0.5182(Affirmative) < succ>
	bacterial outside --- Certainty=0.0000(Not Clear) < succ>
20	bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB11923 GB:Z99104 ybaF [Bacillus subtilis]
 Identities = 133/263 (50%), Positives = 191/263 (72%)

25 Query: 7 MDKLILGRYIPGNSLIHKLDPRSKLLAMLLFIIIVFWANNVVTNVIVFIFTLVIVGLSQI 66
 MD +I+G+Y+PG SL+H+LDPR+KL+ + LF+ IVF ANN T ++ +FT+ +V L+++
 Sbjct: 2 MDSMIIGKYVPGTSLVHRLDPRTKLITIFLFCIVFLANNVQTYALLGLFTIGVVSILTRV 61

30 Query: 67 KFSYFFNGIKPMVGIIIFTTLFQMLFAQGGQVIFSWIFITSLSLGLQQAALIFMRFVLII 126
 FS+ G+KP++ I+LFT L +L G +IF + GL Q I +RFV +I
 Sbjct: 62 PFSFLMKGLKPIIWIVLFTFLHLILMTHEGPPIIFQIGFSRVYEGGLVQGFISLRFVYLI 121

35 Query: 127 FFSSTLLTLTTTSLADAVESLLKPLEVLRVPAHEIGLMLSLSLRFVPTLMDDTTRIMNA 186
 +TLLTLTTT+ + D +E LL PL+ L++P HE+ LM+S+SLRF+PTLM++T +IM A
 Sbjct: 122 LITTLTLTTTPIETDGMQLLNELKLLKLPVHELALMMSISLRFIPTLMBEETDKIMKA 181

40 Query: 187 QRARGVDFGEGNLIHKVKSIIPIILPLFASSFKRADALAIAMEARGYQGGANRKYRLK 246
 Q ARGVDF G + +VK+I+P+L+PLF S+FKRA+ LA+AMEARGYQGG R+KYR L
 Sbjct: 182 QMARGVDFTSGEVFERVKAIVPLLVPLFVSAFKRABELAVAMEARGYQGGEGRTKYRKLIV 241

45 Query: 247 WTVRDTFSILLMLLLGLSLFLK 269
 WT +DT I+ +++L LF L+
 Sbjct: 242 WTKDTSVIVSLIVLAALLFSLR 264

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3175> which encodes the amino acid sequence <SEQ ID 3176>. Analysis of this protein sequence reveals the following:

Possible site: 53
 >>> Seems to have no N-terminal signal sequence

50	INTEGRAL	Likelihood = -9.50	Transmembrane	246 - 262 (243 - 265)
	INTEGRAL	Likelihood = -9.34	Transmembrane	110 - 126 (103 - 135)
	INTEGRAL	Likelihood = -6.69	Transmembrane	41 - 57 (40 - 58)
	INTEGRAL	Likelihood = -2.81	Transmembrane	23 - 39 (21 - 40)
	INTEGRAL	Likelihood = -1.01	Transmembrane	62 - 78 (62 - 78)
55	INTEGRAL	Likelihood = -0.27	Transmembrane	193 - 209 (193 - 209)

----- Final Results -----

	bacterial membrane --- Certainty=0.4800(Affirmative) < succ>
	bacterial outside --- Certainty=0.0000(Not Clear) < succ>
60	bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:CAB11923 GB:Z99104 ybaF [Bacillus subtilis]
 Identities = 138/263 (52%), Positives = 195/263 (73%)

-1144-

Query: 1 MDKLILGRYIPGDSLIHRLDPRSKLLAMIYIVIIIFWANNVVTNLLMLTFTLAVVFLSKI 60
 MD +I+G+Y+PG SL+HRLDPR+KL+ + +++ I+F ANNV T L+ FT+ VV L+++
 Sbjct: 2 MDSMIIGKYVPGTSLVHRLDPRKLITIFLFCIVFLANNVQTYALLGLFTIGVVSILTRV 61

5 Query: 61 KLSFFLNGVKPMIGIILFTTLFQMFSSQGGKVIFSWWFISITDLGLSQAILIFMRFLVII 120
 SF + G+KP+I I+LFT L + + G +IF F + + GL Q I I +RFV +I
 Sbjct: 62 PFSFLMKGLKPIIWIVLFTFLLHILMTHGPIIFQIGFSRVYEGGLVQGFISLRFVYLI 121

10 Query: 121 FFSTLLTLTTTTPSLSDAVESLLKPLTRFKVPAHEIGLMLSLSLRFVPTLMDDTTRIMNA 180
 +TLLTLTTTP+ ++D +E LL PL + K+P HE+ LM+S+SLRF+PTLM++T +IM A
 Sbjct: 122 LITTLTLTTTTPETDGMQLLNPLKKLKLPVHELALMMSISLRFIPTLMEETDKIMKA 181

15 Query: 181 QRARGVDFGEGNLIQVKSIIPILIPLFASSFKRADALAIAMEARGYQGGEGRTKYRQLD 240
 Q ARGVDF G + ++VK+I+P+L+PLF S+FKRA+ LA+AMEARGYQGGEGRTKYR+L
 Sbjct: 182 QMARGVDFTS GPVKERVKAIVPLLVPLFVSAPFKRAELAVAMEARGYQGGEGRTKYRKL 241

Query: 241 WQLKDSLAI GIVSLLGLLLFFLK 263
 W KD+ I + +L LLF L+
 20 Sbjct: 242 WTGKDTSVIVSLIVLAALLFSLR 264

An alignment of the GAS and GBS proteins is shown below.

Identities = 210/263 (79%), Positives = 237/263 (89%)

25 Query: 7 MDKLILGRYIPGNSLIHKLDPKSLAMLLFIIIVFANNVVTNIVVIFITLVIVGLSQI 66
 MDKLILGRYIPG+SLIH+LDPRSKLLAM+++I+I+FWANNVVTN+++ FTL +V LS+I
 Sbjct: 1 MDKLILGRYIPGDSLIHRLDPRSKLLAMIYIVIIIFWANNVVTNLLMLTFTLAVVFLSKI 60

30 Query: 67 KFSYFFNGIKPMVGIIILFTTLFQMLFAQGGQVIFSWWFISITSLGLQQAALIFMRFLVII 126
 K S+F NG+KPM+GIIILFTTLFQ F+QGG+VIFS+W SIT LGL QA LIFMRFLVII
 Sbjct: 61 KLSFFLNGVKPMIGIILFTTLFQMFSSQGGKVIFSWWFISITDLGLSQAILIFMRFLVII 120

35 Query: 127 FFSTLLTLTTTTPSLADAVESLLKPLEVLRVPAHEIGLMLSLSLRFVPTLMDDTTRIMNA 186
 FFSTLLTLTTTTPSL+DAVESLLKPL +VPAHEIGLMLSLSLRFVPTLMDDTTRIMNA
 Sbjct: 121 FFSTLLTLTTTTPSLSDAVESLLKPLTRFKVPAHEIGLMLSLSLRFVPTLMDDTTRIMNA 180

40 Query: 187 QRARGVDFGEGNLIHKVKSIIPIPLIPLFASSFKRADALAIAMEARGYQGGANRSKYRLK 246
 QRARGVDFGEGNLI KVKSIIPILIPLFASSFKRADALAIAMEARGYQGG R+KYR L
 Sbjct: 181 QRARGVDFGEGNLIQVKSIIPILIPLFASSFKRADALAIAMEARGYQGGEGRTKYRQLD 240

Query: 247 WTVRDTFSILLMLLLGLSLFLK 269
 W ++D+ +I ++ LLGL LF LK
 Sbjct: 241 WQLKDSLAI GIVSLLGLLLFFLK 263

45 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1030

A DNA sequence (GBSx1101) was identified in *S. agalactiae* <SEQ ID 3179> which encodes the amino acid sequence <SEQ ID 3180>. This protein is predicted to be unnamed protein product. Analysis of this protein sequence reveals the following:

Possible site: 45

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL Likelihood = -12.05 Transmembrane 22 - 38 (16 - 43)

55 ----- Final Results -----

bacterial membrane --- Certainty=0.5819(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

-1145-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3181> which encodes the amino acid sequence <SEQ ID 3182>. Analysis of this protein sequence reveals the following:

Possible site: 31

5 >>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

10 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 116/233 (49%), Positives = 140/233 (59%), Gaps = 39/233 (16%)

15 Query: 9 KLNVKKHHLAYGAITLVALFSCILAVMVIFKSSQVTTESLSKADKVRVAKKSK----- 61
K N+K+ + +G LVAL ILA++ F S T+S +K + ++ K
Sbjct: 4 KENLKQRYFNFPG---LVALALTILAIIFAFSSKNADTKSYAKKSESKMVTIDKAPKNNHA 60

20 Query: 62 MTKATSKSKVEDVKQAPKPSQASNEAPKSSSQSTEANSQQQVITASEEAAVEQAVVTENTP 121
+TK SK K + + P P+ ++ AP T +EE V Q VT
Sbjct: 61 ITKEESKEKAKSIASEPIPTVENSVAPE-----TVTEVPVQQEVT----- 101

25 Query: 122 ATSQAQQAYAVTETTYRPAQHQTSGQVLSNGNTAGAGSAAAAQMAAATGVPGQSTWEHII 181
Q V+ Y P + VLSNGNTAG +GS AAAQMAAATGVPGQSTWEHII
Sbjct: 102 -----CTVQQVSSVAYNP-----NNVVLNSGNTAGIVGSQAAAQMAAATGVPGQSTWEHII 151

30 Query: 182 ARESNGNPNVANASGASGLFQTMPGWGSTATVQDQVNSAIKAYRAQGLSAWGY 234
ARESNGNPN ANASGASGLFQTMPGWGSTATV+DQVN+A+KAY AQGLSAWGY
Sbjct: 152 ARESNGNPNANASGASGLFQTMPGWGSTATVEDQVNAALKAYSAAQGLSAWGY 204

A related GBS gene <SEQ ID 8713> and protein <SEQ ID 8714> were also identified. Analysis of this protein sequence reveals the following:

35 Lipop: Possible site: -1 Crend: 8
McG: Discrim Score: 2.48
GvH: Signal Score (-7.5): -3.74
Possible site: 45
>>> Seems to have an uncleavable N-term signal seq
ALOM program count: 1 value: -12.05 threshold: 0.0
40 INTEGRAL Likelihood = -12.05 Transmembrane 22 - 38 (16 - 43)
PERIPHERAL Likelihood = 4.29 156
modified ALOM score: 2.91

*** Reasoning Step: 3

45 ----- Final Results -----
bacterial membrane --- Certainty=0.5819(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

50 The protein has homology with the following sequences in the databases:

61.8/68.7% over 114aa

Staphylococcus aureus

GP|7959131| secretory protein SAI-B Insert characterized

55 ORF01057(664 - 1002 of 1302)
GP|7959131|dbj|EAA95959.1||AB042839(119 - 233 of 233) secretory protein SAI-B
{Staphylococcus aureus}
%Match = 15.1
%Identity = 61.7 %Similarity = 68.7
60 Matches = 71 Mismatches = 34 Conservative Sub.s = 8

438 468 498 528 558 588 618 648

```
IFKSSQVTTESLSKADIKVRVAKKSKMTKATSKSKVEDVKQAPKPQSASNEAPKSSSQSTEANSQQQVTASEEAAVEQAIV  
VDQAHLVDLAHNHQDQLNAAPIKDGAydiHFvkdGFqYNFTSNGTtWSwSYEAANGQTAGfSNvagadyTTSYnQGsnVQ  
  
          50           60           70           80           90           100         110  
  
678       708       735       762       792       822       852       882  
TENTPATSQAQQAYAVTETTYRP-AQHQTSGQV-LSNGNTAGAIGSAAAQMAAATGVPQSTWEHI IARESNGNPVNANA  
::|||:||:|||||:|||||:|||||:|||||:|||||:|||||:  
SVSYNAQSSNSNVEAVSAPTYYHNYSTSTSSSVRLSNGNTAGATGSSAAQIMAORTGVFPASTWAAI IARESNQOVNAYNF  
        130      140      150      160      170      180      190  
  
912       942       972       1002     1032     1062     1092     1122  
SGASGLFQTMPGWGSTAtVQDVNSAIKAyRAQGLSAWGy**IAIN*LytVVNNMYRLLkQINKnatVKL*RfyLFsgKE  
|||||:|||||: || |:|:|:|:|:|  
SGASGLFCMPGGWPITNIvdQCINAavKaYaQGLgawgf
```

The purified GBS25-GST fusion product was used to immunise mice (lane 1+2+3 products; 20µg/mouse). The resulting antiserum was used for Western blot (Figure 95B), FACS (Figure 95C), and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

Example 1031

```

Possible site: 61
>>> Seems to have no N-terminal signal sequence
    INTEGRAL      Likelihood = -0.85      Transmembrane  205 - 221 ( 205 - 221)
    INTEGRAL      Likelihood = -0.59      Transmembrane  171 - 187 ( 171 - 187)
    INTEGRAL      Likelihood = -0.53      Transmembrane  226 - 242 ( 226 - 242)

----- Final Results -----
    bacterial membrane --- Certainty=0.1341(Affirmative) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

50

```
Query: 1 MFYTIEELVEQANSQHKGNIAELMIQTEIEMTGRSREEIRYIMSRNLEVMKASVIDGLTP 60
MF ++EL+E ++ I+++MI E+E+T +++E+I M NL VM+A+V GL
Sbjct: 1 MFRNVKELIE-ITKEKQILISDVMIQAQEMEVTEKTKEDIFQQMDHNLVMEAAVQKGLEG 59

Query: 61 SKSISGLTGGDAVKMDQYLQSGKTIISDTTILAAVRNAMAVNELNAKMGLVCATPTAGSAG 120
S +GLTGGDAVK+ Y++SGK+ S I+L AV A+A NE+NA MG+CATPTAGSAG
Sbjct: 60 VTSOTGLGGDAVKLQAYIRSGKSLSGTLLDVA SKAVATNEVNAAMGTICATPTAGSAG 119
```

-1147-

Query: 121 CLPAVISTAIEKLNLTTEEQLDFLFTAGAFGLVIGNNASISGAEGGCQAEVGSASAMAAA 180
 +P + EKLN T E+ + FLFTAGAFG V+ NNASISGA GGCQAEVGSAS MAAA
 Sbjct: 120 VVPGTLFAVKEKLNPTREQMIRFLFTAGAFGFVANNASISGAAGGCQAEVGSASGMAAA 179

5 Query: 181 ALVMAAGGTPFQASQAIAFVIKNMLGLICDPVAGLVEVPCVKRNALGSSFALVAADMALA 240
 A+V AGGTF Q+++A+A +KNMLGL+CDPVAGLVEVPCVKRNA+G+S A++AADMALA
 Sbjct: 180 AIVEMAGGTPEQSAEAMAITLKNMLGLVCDPVAGLVEVPCVKRNAMGASNAMIAADMALA 239

10 Query: 241 GIESQIPVDEVIDAMYQVGSSLPFTAFRETAEGGLAATPTGRRYSKEIFG 289
 GI S+IP DEVIDAMY++G ++PTA RET +GGLAATPTGR K+IFG
 Sbjct: 240 GITSRIPCDEVIDAMYKIGQTMPTALRETGQGGLAATPTGRELEKKIFG 288

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3185> which encodes the amino acid sequence <SEQ ID 3186>. Analysis of this protein sequence reveals the following:

15 Possible site: 55
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -1.12 Transmembrane 196 - 212 (196 - 213)
 INTEGRAL Likelihood = -0.27 Transmembrane 226 - 242 (226 - 242)

20 ----- Final Results -----
 bacterial membrane --- Certainty=0.1447(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

25 The protein has homology with the following sequences in the databases:

>GP:CAB13459 GB:Z99112 similar to L-serine dehydratase [Bacillus subtilis]
 Identities = 173/289 (59%), Positives = 222/289 (75%), Gaps = 1/289 (0%)

30 Query: 1 MFYTIEELVKQADQQFNGNIAELMIATEVEMSGRNREDIIKIMSRNLQVMKAAVTEGLTS 60
 MF ++EL++ ++ I+++MIA E+E++ + +EDI + M NL VM+AAV +GL
 Sbjct: 1 MFRNVKELIEITKEK-QILISDVMI AQEMEVTEKTKEDIFQQMDHNLVMEAAVQKGLEG 59

Query: 61 TKSISGLTGGDAVKMDNYIKGNSLSDTTILNAVRNAIAVNELNAKMGLVCATPTAGSAG 120
 S +GLTGGDAVK+ YI+ G SLS IL+AV A+A NE+NA MG +CATPTAGSAG
 35 Sbjct: 60 VTSQTGLTGGDAVKLQAYIRSGKSLGPLILDVASKAVATNEVNAMGTICATPTAGSAG 119

Query: 121 CLPAVLATAIEKLDLSEKEQLEFLFTAGAFGLVIGNNASISGAEGGCQAEVGSAAAMSAA 180
 +P L EKL+ + ++ + FLFTAGAFG V+ NNASISGA GGCQAEVGSAA+ M+AA
 40 Sbjct: 120 VVPGTLFAVKEKLNPTREQMIRFLFTAGAFGFVANNASISGAAGGCQAEVGSASGMAAA 179

Query: 181 ALVKAAGGTSHQASQAIAFVIKNLLGLVCDPVAGLVEVPCVKRNALGASFALVAADMALA 240
 A+V+ AGGT Q+++A+A +KN+LGLVCDPVAGLVEVPCVKRNA+GAS A++AADMALA
 Sbjct: 180 AIVEMAGGTPEQSAEAMAITLKNMLGLVCDPVAGLVEVPCVKRNAMGASNAMIAADMALA 239

45 Query: 241 DIDSQIPVDEVIDAMYQVGSAMPTAFRETAEGGLAATPTGRRYSVEIFG 289
 I S+IP DEVIDAMY++G MPTA RET +GGLAATPTGR +IFG
 Sbjct: 240 GITSRIPCDEVIDAMYKIGQTMPTALRETGQGGLAATPTGRELEKKIFG 288

An alignment of the GAS and GBS proteins is shown below.

50 Identities = 244/290 (84%), Positives = 273/290 (94%)

Query: 1 MFYTIEELVEQANSQHKGNIAELMIQTEIEMTGRSREEIRYIMSRNLEVMKASVIDGLTP 60
 MFYTIEELV+QA+ Q GNIAELMI TE+EM+GR+RE+I IMSRNL+VMKA+V +GLT
 Sbjct: 1 MFYTIEELVKQADQQFNGNIAELMIATEVEMSGRNREDIIKIMSRNLQVMKAAVTEGLTS 60

55 Query: 61 SKSISGLTGGDAVKMDQYLQSGKTISDTTILAAVRNAMAVNELNAKMGLVCATPTAGSAG 120
 +KSISGLTGGDAVKMD Y++ G ++SDTTIL AVRNA+AVNELNAKMGLVCATPTAGSAG
 Sbjct: 61 TKSISGLTGGDAVKMDNYIKGNSLSDTTILNAVRNAIAVNELNAKMGLVCATPTAGSAG 120

60 Query: 121 CLPAVISTAIEKLNLTTEEQLDFLFTAGAFGLVIGNNASISGAEGGCQAEVGSASAMAAA 180
 CLPAV++TAIEKL+L+E+EQL+FLFTAGAFGLVIGNNASISGAEGGCQAEVGSAA+AM+AA
 Sbjct: 121 CLPAVLATAIEKLDLSEKEQLEFLFTAGAFGLVIGNNASISGAEGGCQAEVGSAAAMSAA 180

Query: 181 ALVMAAGGTPFQASQAIAFVIKNMLGLICDPVAGLVEVPCVKRNALGSSFALVAADMALA 240

-1148-

ALV AAGGT QASQAIAFVIKN+LGL+CDPVAGLVEVPCVKRNALG+SFALVAADMALA
 Sbjct: 181 ALVKAAGGTSHQASQAIAFVIKNLLGLVCDPVAGLVEVPCVKRNALGASFALVAADMALA 240

Query: 241 GIESQIPVDEVIDAMYQVGSLSPTAFRETAEGGLAATPTGRRYSKEIFGE 290
 I+SQIPVDEVIDAMYQVGS++PTAFRETAEGGLAATPTGRRYS EIFGE
 Sbjct: 241 DIDSQIPVDEVIDAMYQVGSAMPTAFRETAEGGLAATPTGRRYSVEIFGE 290

SEQ ID 3184 (GBS358) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 176 (lane 6; MW 35kDa).

- 10 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1032

A DNA sequence (GBSx1104) was identified in *S.agalactiae* <SEQ ID 3187> which encodes the amino acid sequence <SEQ ID 3188>. Analysis of this protein sequence reveals the following:

15 Possible site: 28
 >>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 20 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

25 >GP:BAB06216 GB:AP001515 L-serine dehydratase beta subunit [Bacillus halodurans]
 Identities = 101/216 (46%), Positives = 156/216 (71%), Gaps = 2/216 (0%)

Query: 4 LKFQSVFDIIGPVMIGPSSSHTAGAVRIGKVVSIFGE-PSEVTFHLYNSFAKTYQGHGT 62
 +K+++VFDIIGPVMIGPSSSHTAGA RIG+V ++FG+ P + Y SFA+TY+GHGT
 Sbjct: 1 MKYRTVFDIIGPVMIGPSSSHTAGAARIGRVARTLFGQQPERCDIYFYGSFAETYKGHGT 60

30 Query: 63 DKALVAGILGMDTDNPDIKNSLEIAHQKGIKIYWDILKDSNSPHENTAKITVKNQDRSMS 122
 D A+V GIL DT +P I SL++A +KG+++Y+ +++ + HPNTAK+ ++ G+ +
 Sbjct: 61 DVAIVGGILDFTDFDPRIPRSLQLAKEKGV RVVFHE-EEAITDHPNTAKVVQLQGEDQLE 119

35 Query: 123 ITGVSIGGGNIQVTELNGFSVSLTMNTPTLIIVHQDIPGMIKAVTDILSDFNINIAQMN 182
 + GVSIGGG I++ ELNGF + L+ N P +++VH D G+IA V+++L+ INI M V
 Sbjct: 120 VVGVSIGGGKIEIVELNGFHLKLSGNHPAILVVHTDRFGVIASVSNMLAKHEINIGHMEV 179

40 Query: 183 TRESAGEKAIMIEVDSDRCQAVKKIEAIPHILHN 218
 +R+ G++A+M+IEVD ++++E +P++ V
 Sbjct: 180 SRKEKGKEALMVIEVDQNVDDLQLQLERLEPNIVTV 215

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3189> which encodes the amino acid sequence <SEQ ID 3190>. Analysis of this protein sequence reveals the following:

45 Possible site: 30
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 50 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related sequence was also identified in GAS <SEQ ID 9161> which encodes the amino acid sequence <SEQ ID 9162>. Analysis of this protein sequence reveals the following:

55 Possible site: 28
 >>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

5 bacterial outside --- Certainty= 0.300(Affirmative) < succ>
 bacterial membrane --- Certainty= 0.000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty= 0.000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 187/223 (83%), Positives = 205/223 (91%), Gaps = 1/223 (0%)

10 Query: 1 MKHLKFQSVFDIIGPVMIGPSSSHTAGAVRIGKVVSIFGE-PSEVTFHLYNSFAKTYQG 59
 M KQSVFDIIGPVMIGPSSSHTAGAVRIGKVVSIFG+ P EVTFHLYNSFAKTY+G
 Sbjct: 3 MNTQKFQSVFDIIGPVMIGPSSSHTAGAVRIGKVVSIFGDIPDEVTFHLYNSFAKTYRG 62

15 Query: 60 HGTDKALVAGILGMDTDNPDIKNSLEIAHQKGIKIYWDILKDSN+PHNPNTAKITVKNQDR 119
 HGTDKALVAGI+GM TDNPDIKNSLEIAHQKGIKIYWDILKDSN+PHNPNT KI+VK D+
 Sbjct: 63 HGTDKALVAGIMGMTDNPDIKNSLEIAHQKGIKIYWDILKDSNAPHPNTVKISVKKADK 122

20 Query: 120 SMSITGVSIGGGNIQVTELNQFSVSLTMNTPTLIIVHVDIPGMIKVTDLSDFNINIAQ 179
 ++S+TGVSIGGGNIQVTELNQFSVSL+MNTPT++ VH+DIPGMIKVTDLILS NINIA
 Sbjct: 123 TSVITGVSIGGGNIQVTELNQFSVSLMNTPTIVTVHKDIPGMIKVTDLSSNNINIA 182

25 Query: 180 MNVTRESAGEKAIMIIEVDSRDCQAVKKIEAIPHLNHNFFD 222
 MNVTRESAGEKA MIEVDSR+CQ+A +I IPH++NVNFFD
 Sbjct: 183 MNVTRESAGEKATMIIEVDSRECAANQIAKIPHIYNVNFFD 225

SEQ ID 3188 (GBS151) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 31 (lane 3; MW 50kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 188 (lane 11; MW 25kDa) and in Figure 165 (lane 14-16; MW 25.3kDa).

30 The GBS151-GST fusion product was purified (Figure 198, lane 3; Figure 236, lane 8) and used to immunise mice. The resulting antiserum was used for FACS (Figure 289), which confirmed that the protein is immunoaccessible on GBS bacteria.

35 GBS151L was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 127 (lane 8-10; MW 50kDa). GBS151L was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 127 (lane 11 & 12; MW 25kDa), in Figure 128 (lane 7; MW 25kDa) and in Figure 180 (lane 7; MW 25kDa). Purified GBS151L-His is shown in Figure 232 (lanes 5 & 6) and in Figure 240 (lanes 3 & 4).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

40 **Example 1033**

A DNA sequence (GBSx1105) was identified in *S.galactiae* <SEQ ID 3191> which encodes the amino acid sequence <SEQ ID 3192>. This protein is predicted to be tRNA (5-methylaminomethyl-2-thiouridylate)-methyltransferase (trmU). Analysis of this protein sequence reveals the following:

45 Possible site: 47
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

50 bacterial cytoplasm --- Certainty=0.2208(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

-1150-

A related GBS nucleic acid sequence <SEQ ID 10291> which encodes amino acid sequence <SEQ ID 10292> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```

5      >GP:BAB04980 GB:AP001511
      (5-methylaminomethyl-2-thiouridylate)-methyltran sferase
      [Bacillus halodurans]
      Identities = 250/359 (69%), Positives = 292/359 (80%), Gaps = 6/359 (1%)

10     Query: 32  RVVVGMSGGVDSSVTALLLKEQGYDVIGVFMKNWDDTDEFGVCTATEDYKDVAAVADQIG 91
      RVVVGMSGGVDSSVTALLLKEQGYDVIG+FMKNWDDTDE GVCTATEDY+DV V +Q+G
      Sbjct: 10  RVVVGMSGGVDSSVTALLLKEQGYDVIGIFMKNWDDTDENG VCTATEDYQDVVQVCNQLG 69

      Query: 92  IPYYSVNFEKEYWDRVFEYFLAEYRAGRTPNPDVMCNKEIKFAFLDYAMTLGADYVATG 151
      I YY+VNFEKEYWD+VF YFL EY+AGRTPNPDVMCNKEIKFAFL++A+TLGADYVATG
15     Sbjct: 70  IAYYAVNFEKEYWDKVFTYFLEEYKAGRTPNPDVMCNKEIKFAFLNHALTLGADYVATG 129

      Query: 152 HYAQVTRDENGIVHMLRGADNNKDQTYFLSLSQEQQLKTLFPLGHLQKPEVRRRIABEAG 211
      HYAQV ++ +G ++RG D NKDQTYFL+ LSQ+QL + +FPLGHL+K EVR IAE AG
20     Sbjct: 130 HYAQV-KNVDGQYQLIRGKDPNKDQTYFLNALSQQQLSRVMFPLGHLEKKEVRAIAERAG 188

      Query: 212 LATAKKKDGSTGICFIGEKNKDFLGGYLPAPQPGRMVTVDGRDMGEHAGLMYYTIGQRGGL 271
      LATAKKKDGSTGICFIG+++FK+FL YLPAQPG M T+DG G H GLMYT+GQR GL
      Sbjct: 189 LATAKKKDGSTGICFIGKRDFKEFLSSYLPAPQPGEMQTLDEGEVKGTHDGLMYTIGQRQGL 248

25     Query: 272 GIGGQHGGDNKPFVVGKDLNKLIVGQGFYHDSLMSTSLTASEIHFTTRDMPNEFKLEC 331
      GI GG +PWFV+GK+L KNILYVGQGF+H L S L A ++++ ++ EC
      Sbjct: 249 GI---GSGEPWFVIGKNLEKNILYVGQGFHHPGLYSEGLRAIKVNWILRRESDEPFEC 304

      Query: 332 TAKFRYRQPDQKVTYVYVKGNA-RVVFDLQRAITPGQAVVFYNEQECGGGMIDQAYR 389
      TAKFRYRQPD KVTYV + + A V+F + QRAITPGQAVVFY+ CLGGG ID +
30     Sbjct: 305 TAKFRYRQPDQKVTYVYPQSDGAVEVLFAEPQRAITPGQAVVFYDGDVCLGGGTIDHVLK 363

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3193> which encodes the amino acid sequence <SEQ ID 3194>. Analysis of this protein sequence reveals the following:

```

35     Possible site: 29
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.1691(Affirmative) < succ>
40     bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

      RGD motif: 331-333

```

45 The protein has homology with the following sequences in the databases:

```

      >GP:BAB04980 GB:AP001511
      (5-methylaminomethyl-2-thiouridylate)-methyltran sferase
      [Bacillus halodurans]
      Identities = 255/359 (71%), Positives = 293/359 (81%), Gaps = 6/359 (1%)

50     Query: 14  RVVVGMSGGVDSSVTALLLKEQGYDVIGVFMKNWDDTDEFGVCTATEDYKDVAAVADKIG 73
      RVVVGMSGGVDSSVTALLLKEQGYDVIG+FMKNWDDTDE GVCTATEDY+DV V +++G
      Sbjct: 10  RVVVGMSGGVDSSVTALLLKEQGYDVIGIFMKNWDDTDENG VCTATEDYQDVVQVCNQLG 69

      Query: 74  IPYYSVNFEKEYWDRVFEYFLAEYRAGRTPNPDVMCNKEIKFAFLDYAMTLGADYVATG 133
      I YY+VNFEKEYWD+VF YFL EY+AGRTPNPDVMCNKEIKFAFL++A+TLGADYVATG
55     Sbjct: 70  IAYYAVNFEKEYWDKVFTYFLEEYKAGRTPNPDVMCNKEIKFAFLNHALTLGADYVATG 129

      Query: 134 HYAQVKRDENGIVHMLRGADNGKDQTYFLSLSQEQQLKTLFPLGHLQKSEVREIAERAG 193
      HYAQVK + +G ++RG D KDQTYFL+ LSQ+QL + +FPLGHL+K EVR IAERAG
60     Sbjct: 130 HYAQVK-NVDGQYQLIRGKDPNKDQTYFLNALSQQQLSRVMFPLGHLEKKEVRAIAERAG 188

```


-1151-

Query: 194 LATAKKKDSTGICFIGEKNFKQFLSQYLPQKGRMMTIDGRDMGEHAGLMYYTIGQRGGL 253
 LATAKKKDSTGICFIG+FK+FLS YLPAQ G M T+DG G H GLMYT+GQR GL
 Sbjct: 189 LATAKKKDSTGICFIGKRDFFKFLSSYLPQPGEMQTLTGGEVKGTHDGLMYTIGQRQGL 248

5 Query: 254 GIGGQHGGDNQPFVVGKDLSONILYVGQGFYHEALMSNSLDASVIHFTREMPEEFTFEC 313
 GI GG +PWFV+GK+L +NILYVGQGF+H L S L A +++ + FEC
 Sbjct: 249 GI---GGSGEPWFVIGKNLEKNILYVGQGFHHPGLYSEGLRAIKVNWILRRESDEPFEC 304

10 Query: 314 TAKFRYRQPD SHVAVHVRGDKA-EVVFAPQRAITPGQAVVFYDGECLGGGMIDMAYK 371
 TAKFRYRQPD V V+ + D A EV+FAEPQRAITPGQAVVFYDGE CLGG ID K
 Sbjct: 305 TAKFRYRQPDQKVTVPQSDGAVEVLFAEPQRAITPGQAVVFYDGDVCLGGGTIDHVLK 363

An alignment of the GAS and GBS proteins is shown below.

Identities = 332/377 (88%), Positives = 349/377 (92%)

15 Query: 21 GRILMTDNSNIRVVVGMSGGVDSSVTALLLKEQGYDVIGVFMKNWDDTDEFGVCTATEDY 80
 G MTDNS IRVVVGMSGGVDSSVTALLLKEQGYDVIGVFMKNWDDTDEFGVCTATEDY
 Sbjct: 3 GEFMTDNSKIRVVVGMSGGVDSSVTALLLKEQGYDVIGVFMKNWDDTDEFGVCTATEDY 62

20 Query: 81 KDVAADVADQIGIPYYSVNFKEYWDRVFEYFLAEYRAGRTPNPDVMCNKEIKFKAFLDYA 140
 KDVAADVAD+IGIPYYSVNFKEYWDRVFEYFLAEYRAGRTPNPDVMCNKEIKFKAFLDYA
 Sbjct: 63 KDVAADVADKIGIPYYSVNFKEYWDRVFEYFLAEYRAGRTPNPDVMCNKEIKFKAFLDYA 122

25 Query: 141 MTLGADYVATGHYAQVTRDENGIVHMLRGADNNKDQTYFLSQLSQEQQLQKTLFPLGHLQK 200
 MTLGADYVATGHYAQV RDENG VHMLRGADN KDQTYFLSQLSQEQQLQKTLFPLGHLQK
 Sbjct: 123 MTLGADYVATGHYAQVKRDENGTVHMLRGADNGKDQTYFLSQLSQEQQLQKTLFPLGHLQK 182

30 Query: 201 PEVRRIAEEAGLATAKKKDSTGICFIGEKNFKDFLQYLPQKGRMMTVDGRDMGEHAGL 260
 EVR IAE AGLATAKKKDSTGICFIGEKNFK FL QYLPQK GRMMT+DGRDMGEHAGL
 Sbjct: 183 SEVREIAERAGLATAKKKDSTGICFIGEKNFKQFLSQYLPQKGRMMTIDGRDMGEHAGL 242

35 Query: 261 MYTIGQRGGLGIGGQHGGDNKPFVVGKDLSONILYVGQGFYHDSLMSTSLTASEIHFT 320
 MYTIGQRGGLGIGGQHGGDN+PWFVVGKDLSONILYVGQGFYH++LMS SL AS IHFT
 Sbjct: 243 MYTIGQRGGLGIGGQHGGDNQPFVVGKDLSONILYVGQGFYHEALMSNSLDASVIHFT 302

40 Query: 321 RDMPEFKLECTAKFRYRQPD SKVTVPVKGNOARVVFDDLQRAITPGQAVVFYNEQECLG 380
 R+MP EF ECTAKFRYRQPD S V V+V+G++A VVF + QRAITPGQAVVFY+ +ECLG
 Sbjct: 303 REMPEEFTFECTAKFRYRQPD SHVAVHVRGDKAEVVFAPQRAITPGQAVVFYDGECLG 362

45 Query: 381 GGMIDQAYRDDKICQYI 397
 GGMID AY++ + CQYI
 Sbjct: 363 GGMIDMAYKNGQPCQYI 379

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1034

A DNA sequence (GBSx1106) was identified in *S.agalactiae* <SEQ ID 3195> which encodes the amino acid sequence <SEQ ID 3196>. Analysis of this protein sequence reveals the following:

Possible site: 29
 >>> Seems to have a cleavable N-term signal seq.

INTEGRAL	Likelihood = -12.84	Transmembrane	141 - 157 (134 - 165)
INTEGRAL	Likelihood = -11.78	Transmembrane	40 - 56 (36 - 73)
INTEGRAL	Likelihood = -4.35	Transmembrane	68 - 84 (65 - 86)
INTEGRAL	Likelihood = -3.50	Transmembrane	180 - 196 (175 - 199)

55 ----- Final Results -----

bacterial membrane	---	Certainty=0.6137 (Affirmative)	< succ>
bacterial outside	---	Certainty=0.0000 (Not Clear)	< succ>
bacterial cytoplasm	---	Certainty=0.0000 (Not Clear)	< succ>

60

The protein has homology with the following sequences in the GENPEPT database.

-1152-

>GP:CAB15390 GB:Z99121 similar to hypothetical proteins [Bacillus subtilis]
Identities = 71/202 (35%), Positives = 120/202 (59%), Gaps = 5/202 (2%)

5 Query: 1 MISKFILAFMAFFAIMNPISNLPALFALVADDDQKISRRIAAKGVLLAFVIVFVLSGH 60
M S + F++ FA+ NPI N+P F+ L + IA K +L+F I+ F++ GH
Sbjct: 2 MFSFIVHVFISLFAVSNPIGNVPIFLTLTEGYTAAERKAIARKAAILSFILAAFLVFGH 61

10 Query: 61 LLFNLFGITLAALKISGGILVGIIGYKMINGIHSPNTK-NLEEHKD--DPMNVAVSPLAM 117
L+F LF I + AL+++GGI + I Y ++N S + +EHK+ + +++V+PL++
Sbjct: 62 LIFKLFIDINIHALRVAGGIFIFGIAYNLLNAKESHVQSLHHDEHKESKEKADISVTPLSI 121

15 Query: 118 PLLAGPGTIATAMGLSSG--GLSGKLITILAFAILCVIMYVILISANEITKFLGKNAMTI 175
P++AGPGTIAT M LS+G G+ ++ A + + ++ + I+ LGK M +
Sbjct: 122 PIAGPGTIATVMSLSAGHSIGHYAAVMIGIAAVIALTFLFFHYSAFISSKLGKTEMNV 181

Query: 176 ITKMMGLILMTIGIEMLITGIK 197
IT++MGLIL + + M+ G+K
Sbjct: 182 ITRLMGLILAVVAVGMIGAGLK 203

20 No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8715> and protein <SEQ ID 8716> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 3
McG: Discrim Score: 9.79
25 GvH: Signal Score (-7.5): -1.53
Possible site: 29
>>> Seems to have a cleavable N-term signal seq.
ALOM program count: 4 value: -12.84 threshold: 0.0
30 INTEGRAL Likelihood = -12.84 Transmembrane 141 - 157 (134 - 165)
INTEGRAL Likelihood = -11.78 Transmembrane 40 - 56 (36 - 73)
INTEGRAL Likelihood = -4.35 Transmembrane 68 - 84 (65 - 86)
INTEGRAL Likelihood = -3.50 Transmembrane 180 - 196 (175 - 199)
PERIPHERAL Likelihood = 1.27 110
modified ALOM score: 3.07
35 *** Reasoning Step: 3
----- Final Results -----
bacterial membrane --- Certainty=0.6137(Affirmative) < succ>
40 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

45 ORF00620(301 - 891 of 1209)
OMNI|NT01BS3953(11 - 212 of 220) conserved hypothetical protein
%Match = 15.8
%Identity = 35.5 %Similarity = 61.5
Matches = 71 Mismatches = 74 Conservative Sub.s = 52

50 96 126 156 186 216 246 276 306
VQLSSDIVNLTVKLQFT*KVIKQGLCLMIYNEQSHQVKLLFFIMNKNV*AVG*LIRLIVMIKSVNTFN*HLIIK*GNRMI
VQRLSTRRYMMF
10

55 336 366 396 426 456 486 516 546
SKFILAFMAFFAIMNPISNLPALFALVADDDQKISRRIAAKGVLLAFVIVFVLSGHLLFNLFGITLAALKISGGILVG
| : {::}|: ||| |:| : : || : |:| :| : |||:| || : ||::||:::
SFIVHVFISLFAVSNPIGNVPIFLTLTEGYTAAERKAIARKAAILSFILAAFLVFGHLIFKLFIDINIHALRVAGGIFIF
60 30 40 50 60 70 80 90

576 603 627 657 687 711 741 771
IIGYKMINGIHSPNTK-NLEEHKD--DPMNVAVSPLAMPLLAGPGTIATAMGLSSG--GLSGKLITILAFAILCVIMYVI
| | ::| : : :||: : ::||:|:::||||||| | ||:| : : :| : : : :

-1153-

```

GIAYNLLNAKESHVQSLHHDEHKESKEKADISVTPLSIPIIAGPGTIATVMSLSAGHSGIGHYAAVMIGIAAVIALTFLF
      110      120      130      140      150      160      170

801      831      861      891      921      951      981      1011
5  LISANEITKFLGKNAMTIITKMMGLILMTIGIEMLTIGIKIGFHX*PIPSG*LLKDKC*NKFNXNYDGQSSWNL*VFLT
   : : | : ||| | : || : |||| : : | : |
   FHYSAFISSKLGKTEMNVITRLMGLILAVVAVGMIGAGLKGMFPVLTS
      190      200      210      220

```

- 10 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1035

A DNA sequence (GBSx1107) was identified in *S.agalactiae* <SEQ ID 3197> which encodes the amino acid sequence <SEQ ID 3198>. Analysis of this protein sequence reveals the following:

- 15 Possible site: 17
>>> Seems to have no N-terminal signal sequence
- Final Results -----
- 20 bacterial cytoplasm --- Certainty=0.1747(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10289> which encodes amino acid sequence <SEQ ID 10290> was also identified.

- 25 The protein has homology with the following sequences in the GENPEPT database.

- >GP:AAC45494 GB:U80409 glucose inhibited division protein homolog
GidA [Lactococcus lactis subsp. cremoris]
Identities = 394/524 (75%), Positives = 458/524 (87%), Gaps = 2/524 (0%)
- 30 Query: 13 KTLLATINLEMLAFMPCNPSIGGSAKGIVVREIDALGGEMGNIDKTYIQMKMLNTGKGP 72
KTLT TINL M+AFMPCNPSIGGSAKGIVVREIDALGGEMG+NIDKTYIQMKMLNTGKGP
Sbjct: 12 KTLMTINLNMVAFMPCNPSIGGSAKGIVVREIDALGGEMGNIDKTYIQMKMLNTGKGP 71
- 35 Query: 73 AVRALRAQADKALYAQIMKQTVKEQENLTLRQAMIDEILVEDGK--VVGVRTATNQKFS 130
AVRALRAQADK YA +MK TV QENLTLRQ M++E++D K V+GVRT+T ++ A
Sbjct: 72 AVRALRAQADKDEYAASMKNTVSDQENLTLRQGMVBELILDDEKQKVIGVRTSTGTQYGA 131
- 40 Query: 131 KSVVITGTALRGEIILGDLKYSSGPNNSLASVTLADNLRDLGLEIGRFKGTGTPPRVKAS 190
K+V+ITGTALRGEII+G+LKYSSGPNNSL+S+ LADNLR++G EIGRFKGTGTPPRV AS
Sbjct: 132 KAVIITGTALRGEIIGELKYSSGPNNSLSSIGLADNLRIGFEIGRFKGTGTPPRVLAS 191
- 45 Query: 191 SINYEKTEIQPGDEQPNHFSFMSRDEYITDQVPCWLTYTNTLSHDIINQNLHRAPMFSG 250
SI+Y+KTEIQPGDE PNHFSFMS DEDY+ DQ+PCWLTYT SH I+ NLHRAP+FSG
Sbjct: 192 SIDYDKTEIQPGDEAPNHFSFMSDEYLDQDQPCWLTYTTENSHTILRDNLHRAPLFSG 251
- 50 Query: 251 IVKGVGPYPCPSIEDKIVRFADKERHQLFLEPEGRYTEEVYVQGLSTSLPEDVQVDLLRS 310
IVKGVGPYPCPSIEDKI RFADK RHQLFLEPEGR TEEVY+ GLSTS+PEDVQ DL++S
Sbjct: 252 IVKGVGPYPCPSIEDKITRFADKPRHQLFLEPEGRNTEEVYIGGLSTSMPEDEVQFDLVKS 311
- 55 Query: 311 IKGLENAEMMRTGYAIEYDIVLPHQLRATLETKVIAGLFTAGQTNGTSGYEEAAGQGLVA 370
I GLENA+MMR GYAIEYD+V+PHQLR TLETK+I+GLFTAGQTNGTSGYEEAAGQGLVA
Sbjct: 312 IPGLENAKMMRPGYAIEYDVVMPHQLRPTLETKLISGLFTAGQTNGTSGYEEAAGQGLVA 371
- Query: 371 GINAALKVQKPELILKRSDAYIGVMIDDLVTKGTLEPYRLITSRAEYRLILRHDNADMR 430
GINAALK+QKPE ILKRS+AYIGVMIDDLVTKGTLEPYRLITSRAEYRLILRHDNAD R
Sbjct: 372 GINAALKIQKPEFILKRSEAYIGVMIDDLVTKGTLEPYRLITSRAEYRLILRHDNADRR 431
- Query: 431 LTEIGYEIGLVDEERYAIFKKRQMQFENELERLDSIKLKPVSSETNKRIGELGFKPLTDAL 490
LTEIG ++GLV + ++ ++ + QF+ E++RL+S KLKP+ +T +++ +LGF P+ DAL
Sbjct: 432 LTEIGRQVGLVSDAQWEHYQAKMAQFDREMKRLNSEKLKPLPDTQEKLGKLGFGPIKDAL 491

-1154-

Query: 491 TAKEFMRRPQITYAVATDFVGCADPEPLDSKVIELLETEIKYEGY 534
 T EF++RF++ Y DF+G A E +D V EL+ETEI YEGY
 Sbjct: 492 TGAEFLKRPEVNYDEVIDFIGQAPEVIDRTVSELIETETITYEGY 535

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3199> which encodes the amino acid sequence <SEQ ID 3200>. Analysis of this protein sequence reveals the following:

Possible site: 28
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1064(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 530/610 (86%), Positives = 574/610 (93%)

Query: 1 MEASLAASRMGCKTLLATINLEMLAFMPCNPISIGGSARGIVVREIDALGGEMGKNIDKTY 60
 +EASLA SRMGCKTLLATINL+MLAFMPCNPISIGGSARGIVVREIDALGGEMGKNIDKTY
 Sbjct: 21 VEASLATSRMGCKTLLATINLMDLAFMPCNPISIGGSARGIVVREIDALGGEMGKNIDKTY 80

Query: 61 IQMKMLNTGKGPVAVRALRAQADKALYAQTMKQTVKQENLTLRQAMIDEILVEDGKVVGV 120
 IQMKMLNTGKGPVAVRALRAQADK+LYA+ MK TVEKQ NLTLRQ MID+ILVEDG+VVG
 Sbjct: 81 IQMKMLNTGKGPVAVRALRAQADKSLYAREMKHTVEKQANLTLRQTMIDDILVEDGRVGV 140

Query: 121 RTATNQKFSAKSVVITTTGTALRGEIILGDLKYSSGPNNSLASVTLADNLRDLGLEIGRFK 180
 TAT QKF+AK+VV+TTGTALRGEIILG+LKYSSGPNNSLASVTLADNL+ LGLEIGRFK
 Sbjct: 141 LTATGCKFAAKAVVVTTGTALRGEIILGELKYSSGPNNSLASVTLADNLKGLGLEIGRFK 200

Query: 181 TGTTPRVKASSINYEKTEIQPGDEQPNHFSFMSRDEYITDQVPCWLTYTNTLSHDIINQ 240
 TGTTPRVKASSINY++TEIQPGD++PNHFSFMS+D DY+ DQ+PCWLTYTN SHDIINQ
 Sbjct: 201 TGTTPRVKASSINYDQTEIQGDDKPNHFSFMSKDADYLKDQIPCWLTYNQTSHDIINQ 260

Query: 241 NLHRAPMFGSIVKGVGPYRCPSEDKIVRFADKERHQLFLEPEGRYTEEVYVQGLSTSLP 300
 NL+RAPMFGSIVKGVGPYRCPSEDKIVRFADKERHQLFLEPEGR TEEVYVQGLSTSLP
 Sbjct: 261 NLYRAPMFGSIVKGVGPYRCPSEDKIVRFADKERHQLFLEPEGRDTEEVYVQGLSTSLP 320

Query: 301 EDVQVDLLRSIKGLENAEMMRTGYAIEYDIVLPHQLRATLETKVIAGLFTAGQTNGTSGY 360
 EDVQ DL+ SIKGLE AEMMRTGYAIEYDIVLPHQLRATLETK+I+GLFTAGQTNGTSGY
 Sbjct: 321 EDVQKDLIHSIKGLEKAEMMRTGYAIEYDIVLPHQLRATLETKLISGLFTAGQTNGTSGY 380

Query: 361 EEAAGQGLVAGINAALKVQGKPELILKRSDAYIGVMIDDLVTKGTLEPYRLTTSRAEYRL 420
 EEAAGQGL+AGINAALKVQGKPELILKRSDAYIGVMIDDLVTKGTLEPYRLTTSRAEYRL
 Sbjct: 381 EEAAGQGLIAGINAALKVQGKPELILKRSDAYIGVMIDDLVTKGTLEPYRLTTSRAEYRL 440

Query: 421 ILRHDNADMRLTEIGYEIGLVDEERYAIFKKRQMQFENELERLDSIKLKPVSETNKRIQE 480
 ILRHDNADMRLTEIG +IGLVD+ER+ F+ ++ QF+NEL+RL+SIKLP+ ETN R+Q+
 Sbjct: 441 ILRHDNADMRLTEIGRDIGLVDDERWKAFEIKKNQFDELKRLNSIKLKPiketNDRVQD 500

Query: 481 LGFKPLTDALTAKEFMRRPQITYAVATDFVGCADPEPLDSKVIELLETEIKYEGYIKKALD 540
 LGFKPLTDA+TAKEFMRRP+I YA A FVG A E LD+K+IELLETEIKYEGYI+KALD
 Sbjct: 501 LGFKPLTDAMTAKEFMRRPEIDYATAVSFVGPAAEDLDAKIELLETEIKYEGYIRKALD 560

Query: 541 QVAKMKRMEEKRIPPHIDWDDISIAEARQKFKKINPETLGQASRISGVNPADISILMV 600
 QVAKMKRMEEKRIP +IDWD IDSIATEARQKFKKINPET+GQASRISGVNPADISILM+
 Sbjct: 561 QVAKMKRMEEKRIPTNIDWDAIDSIATEARQKFKKINPETIGQASRISGVNPADISILMI 620

Query: 601 YLEGRQKGRK 610
 YLEG K +
 Sbjct: 621 YLEGNGKAHR 630

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1155-

Example 1036

A DNA sequence (GBSx1108) was identified in *S. agalactiae* <SEQ ID 3201> which encodes the amino acid sequence <SEQ ID 3202>. Analysis of this protein sequence reveals the following:

Possible site: 44
 5 >>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 10 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:BA07750 GB:AP001520 unknown conserved protein in B. subtilis
 [Bacillus halodurans]
 15 Identities = 205/644 (31%), Positives = 362/644 (55%), Gaps = 28/644 (4%)

Query: 35 LLLAIFVALSFVVALLYQ-----KITVELSEVEQIELLNDQTE 73
 ++ + VAL F++AL +YQ +I++E + I L+ +
 20 Sbjct: 14 VIALLAVALVFLIALSFYQWQLGVIGVLLLVIAIFSLRARISFERDLEQYISTLSYRVH 73

Query: 74 VSLKSLLEQMPVGVQFDLENDIEWFNPYA-ELIFTGDNGHFQSATVKDIITSRRNGTA 132
 + + + Q+PVG+I ++ + ++W NPYA E + + +++ + GT
 25 Sbjct: 74 KAGEEAVTQLPVGMIYNDQLR-VQWVNPYAAEHLPKAEIDASLEELSPELVRALEEGTD 132

Query: 133 GQSFYEGDNKYSAYLDTETGVFYFFDNFMGNRRNYDSSMLRPVIGIISIDNYDDIMDTML 192
 Q + Y + YFFD R + +PV+ I +DNYD++ M
 30 Sbjct: 133 EQKIVIEEKTYDCTFKPNERLIYFFDITESERMHQFEESQPVLTFFIYLDNYDEVTDQME 192

Query: 193 EADMSKINAFVTSFISDFTQSKNIFYRRVNMDDRYIIFTDYSVLNTLIKDKFDILNEFRKR 252
 + S++ + VTS ++ + ++F RR DR+ Y L + K KF IL+E R+
 35 Sbjct: 193 DQVRSLMSQVTSLSLQWANEHDLFLRRTAADRFAVMSYGSLLAEKTKFGILDEIRET 252

Query: 253 AQENHLSLTLSMGISYGDGNHNQIGQIALENLNTALVRGGDQIVRENDSSKALYFGGG 312
 + + LTLG+G+ YGD + ++GQ+A +L+ AL RGGDQ+ +++ K ++GG
 40 Sbjct: 253 TGKEKIPLTLSIGVGYGDLRLRELQQLAQSSLDLALGRGGDQVAIKQKTG--KVRFYCGK 310

Query: 313 AVSTIKRSRTRTRAMMTAISDRKLVVDSVFIVGHRKLDMDALGASVGMQFFASNIVNASY 372
 + + KR+R R R + A+ D + D V ++GH+ DMDA+GA++G+ A ++
 45 Sbjct: 311 SNAMEKRTVRARVISHALRFVLES DRVIVMGHKNPMDMDAVGAAGILKIAEVNDREAF 370

Query: 373 VVYDPNDMNSDIERAIDYQLQEDGET--RLVSVERAFELITQNSLLVMVDHSTALTLSKE 430
 VV DPND+N D+ + ++ +++++ + + ++ E + EL+T+ +LLV+VD K ++ +
 50 Sbjct: 371 VVLDPNVDNPDVSKLMEEVEKNEQLWDFITPEESLELMTEETLLVIVDTHKPSMVIEPR 430

Query: 431 FPNKPADVIVVDHHRDEDFFPKNAVLSFIESGASSASELVTELIQFQAKDKLSRSQASI 490
 + V+V+DHRR E+F ++ VL ++E ASS +ELVTEL+++Q K K+ +++
 55 Sbjct: 431 LLDYVERVVLDHHRGEEFIEDPVLVMEFYASSTAELVTELLLEYQPKLKMIDLESTA 490

Query: 491 LMAGIMLDTRNFASNVTSRTFDVASYLRLGLGSNSMAIQKISATDFDEYRLINELILKGER 550
 L+AG+++DT++FA +RTFD AS+LR G++++ +OK+ D + Y +L+ +
 60 Sbjct: 491 LLAGMIVDTKSFAIRTGARTFDAASFLRSHGADTVLVQKLLKEDLNHYVKRAKLVETAKL 550

Query: 551 IYDNIIVATGEEHKVYSHVIASKAADTMLTMAGIEATFVITKNSSN-IGISARSRNNINV 609
 D + +AT E + S ++ ++AADT+LTM G+ A+FVI++ + ISARS ++NV
 65 Sbjct: 551 YRDGMAIATAREEEAVSOLLIAQAADTLLTMKGVVASFVISRRHDGVVISARSLGDVNV 610

Query: 610 QRIMEKLGGGGHFSFAACQIQDKSVKQVRRMLLEIIDEIDRENS 653
 Q IME L GGGH + AA Q +D +++++ L E ID+ L S
 70 Sbjct: 611 QLIMESLDGGGHLTNAATQFEDATLEEAELKEAIDQYLEGGS 654

A related DNA sequence was identified in *S. pyogenes* <SEQ ID 3203> which encodes the amino acid sequence <SEQ ID 3204>. Analysis of this protein sequence reveals the following:

Possible site: 25

-1156-

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL Likelihood = -18.57 Transmembrane 33 - 49 (6 - 56)
 INTEGRAL Likelihood = -10.14 Transmembrane 12 - 28 (6 - 32)

5

----- Final Results -----

bacterial membrane --- Certainty=0.8429(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

10 The protein has homology with the following sequences in the databases:

>GP:BA07750 GB:AP001520 unknown conserved protein in B. subtilis

[Bacillus halodurans]

Identities = 199/659 (30%), Positives = 367/659 (55%), Gaps = 16/659 (2%)

15

Query: 1 MKKF---RFETIHLI-MMGLILFGLLALCVSIMQSKILILLAIPLVLLFVV-ALLWYQKE 55
 M KF R+ H+I ++ + L L+AL Q ++ +L + ++ +F + A + +++
 Sbjct: 1 MPKFLLRWGHYHVIALLAVALVFLIALSFYQWQLGVIGVLLLVIAIFSLRARISFERD 60

20

Query: 56 AYQLSDLAHIELLNEQTEDNLKTLNMPVGVVQFDQETNAVEWYNPYA-ELIFTTEEGF 114
 Q +I L+ + + + +PVG++ ++ + V+W NPYA E + E
 Sbjct: 61 LEQ-----YISTLSYRVHKAGEBAVTQLPVGMIYNDQLR-VQWVNPYAAEHLPAEIDA 114

25

Query: 115 IQNGLIQQIITEKRREDISQTFEVSQNKYTSYIDVSSGIFYFFDSFVGNRQLADASMLRP 174
 L +++ Q + Y + + YFFD R +P
 Sbjct: 115 SLEELSPELVRALLEGTEQKIVIEEKTYDCTFKPNERLIYFFDITESERMHQQFEESQP 174

30

Query: 175 VVGIIISVDNYDDITDLSADTSKINSFVANFIDEFMESKRIFYRRVNMDDRYFFTFDKT 234
 V+ I +DNYD++T + D S++ S V + +++ +F RR DR+ + +
 Sbjct: 175 VLTFIYLDNYDEVTOGMEQVRSRLMSQVTSSLNQWANEHDLFLRRTAADRFAVMSYGS 234

35

Query: 235 LNDLMDNKFVLEEFKREQAQDAQRPLTSLSIGISFGEENHSQIGQVALENLIALVRGGDQ 294
 L + KF +L+E R+ + PLTSLIG+ +G+ + ++GQ+A +L++AL RGGDQ
 Sbjct: 235 LLAIEKTKFGILDEIRETTGKEKIPLTSLSIGVGYDLSRELQQLAQSSLDLALGRGGDQ 294

40

Query: 295 IVIRENADHTNPIYFGGGSVSTVKRSRTRTRAMTAISDRIKMVDNVFIVGHRKLDMDAL 354
 + I++ ++GG S + KR+R R R + A+ D + D V ++GH+ DMDA+
 Sbjct: 295 VAIKQKTGKVR--FYGGKSNAMEKRTFRVARVISHALRDFVLES DRVIVMCHKNPMDMAV 352

45

Query: 355 GSAVGMQFFAGNIIEENSFAVYNPDMSPDIERAIERLQADGKT--RLISVSQAMGLVTPR 412
 G+A+G+ A +F V +P++++PD+ + +E ++ + + I+ +++ L+T
 Sbjct: 353 GAAIGILKIAEVNDREAFVVLDPNDVNPVSKLMEEVEKNEQLWDFITPEESLELMTEE 412

50

Query: 413 SLLVMVDHKSISLTLSKEFYEQFNQVIVVDHHRDDDFPDNAILTFIESGASSAAELVTE 472
 +LLV+VD K S+ + + + V+V+DHRR ++F ++ +L ++E ASS AELVTE
 Sbjct: 413 TLLVIVDTHKPSMVIEPRLLDYVERVVDHHRGEEFIEDPVLVMEPYASSTAELVTE 472

55

Query: 473 LIQFQNAKKCLNKIQASVLMAGIMLDTKNFSTRVTSRTFDVASYLRSKGSDSVEIQNISA 532
 L+++Q K ++ +++ L+AG+++DTK+F+ R +RTFD AS+LRS G+D+V +Q +
 Sbjct: 473 LLEYQPKKLKMDILESTALLAGMIVDTKSFARTGARTFDAASFLRSHGADTVLVQKLLK 532

Query: 533 TDFEEYKQINEIILQGERLGSIIVAAGEKNHLYSNVIASKAADTILSMARVEASFVLVE 592
 D Y + +++ + D + +A + S ++ ++AADT+L+M V ASFV+
 Sbjct: 533 EDLNHYVVKRAKLIVETAKLYRDGMAIATAREEEAVSQLLIAQAADTLLTMKGVVASFVVISR 592

Query: 593 TASHKIAISARSRSKINQVRVMEKLGCGGHFNLAACQLTDISLPQAKYLLKKTINMTMK 651
 ++ISARS +NVQ +ME L GCGH AA Q D +L +A+ L + I+ ++
 Sbjct: 593 RHDGVVISARSRLGDVNVQLIMESLDGCGHLTNAATQFEDATLEEAELKRAIDQYLE 651

An alignment of the GAS and GBS proteins is shown below.

60

Identities = 428/658 (65%), Positives = 547/658 (83%), Gaps = 1/658 (0%)

Query: 1 MKRFRFATVHLVLIGLILFGLLAICVRLFQSYTALLLAIFVALSFVVALLYQKITYELS 60
 MK+FRF T+HL+++GLILFGLLA+CV + QS +LLAIF+ L FVVALL+YQK Y+LS
 Sbjct: 1 MKKFRFETIHLIMGLILFGLLALCVSIMQSKILILLAIPLVLLFVVALLWYQKRAYQLS 60

65

Query: 61 EVEQIBLLNDQTEVSLKSLLEQMPVGVVQFDLETDNDIEWFNFPYAEIIFTGDNHGRQSATV 120

-1157-

++ IELLN+QTE +LK+LL+ MPVGV+QFD ETN +EW+NPYAE LIFT + G Q+ +
 Sbjct: 61 DLAHIELLNEQTEBDNLKTLDDNMPVGVVQFDQETNAVEWYNPYAE LIFTTEEGFIQNGLI 120

Query: 121 KDIITSRRNGTAGQSFYEGDNKYSAYLDTETGVFYFFDNFMGNRRNYDSSMLRPVIGIIS 180
 + IIT +R Q+FE NKY++Y+D +G+FYFFD+F+GNNR+ D+SMLRPV+GIIS
 Sbjct: 121 QQIITEKRREDISQTFEVSGNKYTSYIDVSSGIFYFFDSFVGNRQLADASMLRPVVGIIIS 180

Query: 181 IDNYDDIMDTMLEADMSKINAFVTSFISDFTQSKNIFYRRVNMDDRYIIFTDYSVLNLTLIK 240
 +DNYDDI D + +AD SKIN+FV +FI +F +SK IFYRRVNMDDRYI FTD+ LN L+
 Sbjct: 181 VDNYYDDITDDLSADATSKINSFVANFIDEFMESKRIFYRRVNMDDRYI FTD+ LN L+ 240

Query: 241 DKFDILNEFRKRAQENHLSLTLSMGISYGDGNHNOIGQIALENLNTALVRGGDQIVVREN 300
 +KF +L EFRK AQ+ LTL+GIS+G+ NH+QIGQ+ALENLN ALVRGGDQIV+REN
 Sbjct: 241 NKFSVLEEFKRAQDAQRPLTSLGISFGEENHSQIGQVALENLNTALVRGGDQIVIREN 300

Query: 301 DSSKKALYFGGGAVSTIKRSRTRTRAMMTAISDRKLVDSVFIVGHRKLDMDALGASVGM 360
 +YFCGG+VST+KRSRTRTRAMMTAISDR+K+VD+VFIVGHRKLDMDALG++VGM
 Sbjct: 301 ADHTNPIYFGGGSVSTVKRSRTRTRAMMTAISDRKLVDSVFIVGHRKLDMDALGSAVGM 360

Query: 361 QFFASNIVNASYVVYDPNDMNSDIERAIDYLQEDGETRLVSVERAFELITQNSLLVMVDH 420
 QFFA NI+ S+ VY+P++M+ DIERAI+ LQ DG+TRL+SV +A L+T SLLVMVDH
 Sbjct: 361 QFFAGNIIENSFAVYNPDMSPIERIAERLQADGKTRLISVSQAMGLVTPRSLLVMVDH 420

Query: 421 SKTALTLSKEFFNKFADVIVVDHHRDEDFPKNAVLFSFIESGASSASELVTELIQFQQA 480
 SK +LTLKEF+ +F +VIVVDHHRD+DFP NA+L+FIESGASSA+ELVTELIQFQ AK
 Sbjct: 421 SKISLTLSKEFYEQFQNVIVVDHHRDDDFPDNAILTFIESGASSAELVTELIQFQNAK 480

Query: 481 DKLSRSQASILMAGIMLDTRNFASNVTSRTFDVASYLRLGSGNSMAIQKISATDFDEYRL 540
 L++ QAS+LMAGIMLDT+NF++ VTSRTFDVASYLRL GS+S+ IQ ISATDF+EY+
 Sbjct: 481 KCLNKIQASVLMAGIMLDTKNFSTRVTSRTFDVASYLRLSGSDSVEIQNISATDFEYKQ 540

Query: 541 INELILKGERIYDNIIVATGEEHKVYSHVIASKAADTMLTMAGIEATFVITKNSSN-IGI 599
 INE+IL+GER+ D+IIVA GE++ +YS+VIASKAADT+L+MA +EA+FV+ + +S+ I I
 Sbjct: 541 INEIIQGERLGDSEIIVAAGEKNHLYSNVIASKAADTILSMHVEASFVIVETASHKIAI 600

Query: 600 SRSRNNINVQRIMEKLGCGGHFSFAACQIQDKSVKQVRMLLEIIDEEDLRENSTVEN 657
 SRSR+ INVQR+MEKLGCGGHF+ AACQ+ D S+ Q + +LL+ I+ ++E VE+
 Sbjct: 601 SRSRSKINVQRVMEKLGCGGHFNLAACQLTDISLPQAKYLLKTINMTMKETGEVES 658

40 A related GBS gene <SEQ ID 8717> and protein <SEQ ID 8718> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 9
 McG: Discrim Score: 13.82
 GvH: Signal Score (-7.5): -0.890001
 45 Possible site: 44
 >>> Seems to have a cleavable N-term signal seq.
 ALOM program count: 0 value: 2.97 threshold: 0.0
 PERIPHERAL Likelihood = 2.97 574
 modified ALOM score: -1.09
 50 *** Reasoning Step: 3
 ----- Final Results -----
 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 55 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

31.3/55.8% over 631aa
 Bacillus subtilis
 60 EGAD|19304| hypothetical 74.3 kd protein in rpli-cotf intergenic region Insert
 characterized
 SP|P37484|YYBT_BACSU HYPOTHETICAL 74.3 KDA PROTEIN IN RPLI-COTF INTERGENIC REGION. Insert
 characterized
 GP|467336|dbj|BAA05182.1||D26185 unknown Insert characterized
 65 GP|2636598|emb|CAB16088.1||Z99124 yybT Insert characterized

EGAD|19304|BS4045(20 - 651 of 659) hypothetical 74.3 kd protein in rpli-cotf intergenic region {*Bacillus subtilis*}SP|P37484|YYBT BAC

98|emb|CAB16088.1||Z99124 yybT {Bacillus subtilis}PIR|S65976|S65976 yybT protein - Bacillus subtilis

%Match = 18.5

%Identity = 31.2 %Similarity = 55.8

Matches = 197 Mismatches = 271 Conservative Sub.s = 155

258 288 318 348 378 408 438 468
N***CSPLFIRGVLCYN*VLRGYLMKFRFRFATVHLVLIGLILGFLLAICVRLQSYTALLAIFVALSFVVALLYQKIT
| | : : |:| | | : | : | :::
MPSFYEKLPFRYPYIALIASIITILISFYFNWILGTVEVLLAVILFFIKRAD
10 20 30 40 50

522 552 582 612 666 696
YEL-SEVEQ-IELLNDQTEVSLKSLLEQMPVGVIOFDLETNDBWFNPYAEILFTGDN--GHFQSATVKDIIITSRRNGTA
: |:: | |: : : : |::|::|: : ||| |: | | : | : : :
SLIRQEBIDAYISTLSYRLKKVGEELMEMPIGIMLFN-DQYYIEWANPFLSSCFNESTLVGRSLYDTCESVVP LIKQVEE
70 80 90 100 110 120 130

726 756 786 816 846 876 906 936
 GQSFEYGDNNKYSAYLDTETGVFFFFDNFMGNRRNYDSSMLRPVIGIISIDNYDDIMDTMLEADMSKINAXVTSFXSDFTO
 :: | | : : : : || : : | : | : || | : : | : : |
 SETVTLNDRKFRVVIKRDERLLYFFDVTEQIQIEKLYENERTVLAYIFLDNYDDVTOGLDDQTRSTMNSQVTSLLNNAWAQ

150 160 170 180 190 200 210

966 996 1026 1056 1086 1116 1146 1176
SKNIFYRRVNMDDRYIYFTDYSVLNTLIKDKFDILNEFRKRAQENHLSITLMSGISYGDNHNHQIGQIALENLNTALVRGG
|| :| : :| : :| | || |:| :: : :||| :| : :| :| :| :|
EYGIPLKRTSSERFI AVLNEHILTELENSKFSILDEVREKTSFDGVALTL SVGVGASVSSSLKELGDLAQQSSLDLALGRGG
230 240 250 260 270 280 290

1206 1236 1266 1296 1326 1356 1386 1416
DQIVVRENDSSKKALYFGGGAVSTIKRSRTRTRAMMTAISDRLKVVDSVFIVGHRKLLMDALGASVGMQFFASNIVNASY
||: :: : | ::|| ||:| ||: |::: :| |::|||::|::|: |
DQVAIKLPGNKVKV--FYGGKTNPMEKRTVRARVISHALKEIVTESSNVIIMGHKFPDMSDIGAIGILKVAQANNKDG
 310 320 330 340 350 360 370

VVYDPNDMNSDIERAIDYLQEDGE--TRLVSVERAFELITQNSLLVMVDHSKTALTLSKEFFNKFADVIIVDDHHRRDEDF
:
:| || : | :: | :: | :||: | | : :|||:| | : : : | | :::||||| |:|
IVIDPNQIGSSVQRLIGEIKKYELWSRFITPEEAMEISNDTLVLIVDTHKPSLVMEERLVNKIEHVIVIDHHRRGEFF

1446 1476 1500 1530 1560 1590 1620 1650

390 410 420 430 440 450

1680 1710 1740 1770 1800 1830 1860 1890
PKNAVLSTFIESGASSASELVTELIQFQQAKDKLSRSQASILMAGIMLDTRNFASNVTSRTPFDVASYLRLGLGSNSMAIQKI
:: :| ::| ||| :||| |::| :| :: :| :| :||| :| ::| :| ||| | ||| | :| :| :|
IRDPLLVMEFPYASSTAELVTELLEYQPKRLKINMIETALLAGIIVDTKFSRLRTSGSRTFDAASYLRAKGADITVLVKQF
 470 480 490 500 510 520 530

1920 1950 2004 2034 2064 2091 2121
SATDFDEYRLINELILKGERIYDNIIVAT--GEEHKVYSHVIAKAAADTMLIMAGIEATFVITK-NSSNIGISARSRNNI
| | :|| ||| :| : : : : :||| : : :||| : : : : |||| :
LKETVDSYIKRAKLIQHTVLKYDNIAIASLPENEBEYFDQVLIQAADSLLSMSEVEASFVARDEQTVCTISARSLGEV
550 560 570 580 590 600 610

2151 2181 2211 2241 2271 2301 2331 2361
 NVQRIMEKLGGGGHFSFAACIQDKSVKQVRMLLEIIDEDLRENSTVENRRD*LR*KLFFYKMLRGKEKKVRLRKYLLV
 ||| ||| | |||:: || : || : | ||| :
 NVQIIMEALEGGGHLTNAATQLSGISVSEALERLKAIDEIFEGGVQR
 630 640 650

SEQ ID 8718 (GBS10) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 1 (lane 6; MW 98kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 2 (lane 7; MW 73kDa).

The GST-fusion protein was purified as shown in Figure 189, lane 3.

- 5 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1037

A DNA sequence (GBSx1109) was identified in *S.agalactiae* <SEQ ID 3205> which encodes the amino acid sequence <SEQ ID 3206>. Analysis of this protein sequence reveals the following:

10 Possible site: 59
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
15 bacterial cytoplasm --- Certainty=0.4643(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

20 >GP:CAA43972 GB:X62002 ribosomal protein L9 [Bacillus
stearothermophilus]
Identities = 80/149 (53%), Positives = 105/149 (69%), Gaps = 2/149 (1%)

Query: 1 MKVIFLQDVKGKGGKGEVKEVPTGYAQNFLKKNLAKEATTQAIGELKGGKQKSEBKAQAE 60
MKVIFL+DVKGKGGKGE+K V GYA NFL K+ LA EAT + L+ +++ E++ AE
25 Sbjct: 1 MKVIFLKDVGKGGKGEIKNVADGYANNFLFKQGLAIEATPANLKALEAQKQKEQRQAAE 60

Query: 61 ILAQAKELKTQLESETTRVQFIEKVGPDGRTFGSITAKKIAEELQKQYGIKIDKRHIDL 120
LA AK+LK QLE T + K G GR FGSIT+K+IAE LQ Q+G+K+DKR I+L
30 Sbjct: 61 ELANAKKLKEQLEKLTVTIP--AKAGEGGRFLFGSITSKQIAESLQAQHGLKLDKRKIELA 118

Query: 121 HTIRAIGKVEVPVKLHKQVSSQIKLDIKE 149
IRA+G VPVKLH +V++ +K+ + E
Sbjct: 119 DAIRALGYTNVPVKLHPEVTATLKVHVTE 147

- 35 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3207> which encodes the amino acid sequence <SEQ ID 3208>. Analysis of this protein sequence reveals the following:

Possible site: 59
>>> Seems to have no N-terminal signal sequence

40 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.4630(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 45 An alignment of the GAS and GBS proteins is shown below.

Identities = 119/150 (79%), Positives = 138/150 (91%)

Query: 1 MKVIFLQDVKGKGGKGEVKEVPTGYAQNFLKKNLAKEATTQAIGELKGGKQKSEBKAQAE 60
MKVIFL DVKGKGGKGE+KEVPTGYAQNFL+KKNLAKEAT+Q+IGELKGGKQ+EEKAQAE
50 Sbjct: 1 MKVIFLADVGKGGKGEIKNVADGYANNFLFKQGLAIEATPANLKALEAQKQKEQRQAAE 60

Query: 61 ILAQAKELKTQLESETTRVQFIEKVGPDGRTFGSITAKKIAEELQKQYGIKIDKRHIDL 120
ILA+A+ +K L+ + TRVQF EKVGPDPGRTFGSITAKKI+BELQKQ+G+K+DKRHI LD
Sbjct: 61 ILAQAQAVKAVLDEDKTRVQFQEKVGPDGRTFGSITAKKISELQKQFGVKVDDRHI VLD 120
55

-1160-

Query: 121 HTIRAIGKVEVPVKLHKQVSSQIKLDIKEA 150
 H IRAIG +EVPVKLHK+V+++IKL I EA
 Sbjct: 121 HPIRAIGLIEVPVKLHKEVTAEIKLAITEA 150

- 5 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1038

- A DNA sequence (GBSx1110) was identified in *S.agalactiae* <SEQ ID 3209> which encodes the amino acid sequence <SEQ ID 3210>. This protein is predicted to be DNA polymerase III delta prime subunit (dnaB). Analysis of this protein sequence reveals the following:

Possible site: 61
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -0.43 Transmembrane 204 - 220 (204 - 220)

- 15 ----- Final Results -----
 bacterial membrane --- Certainty=0.1171(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

- 20 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2423> which encodes the amino acid sequence <SEQ ID 2424>. Analysis of this protein sequence reveals the following:

Possible site: 21
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.27 Transmembrane 210 - 226 (210 - 226)

- 25 ----- Final Results -----
 bacterial membrane --- Certainty=0.1107(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

- 30 An alignment of the GAS and GBS proteins is shown below.

Identities = 397/450 (88%), Positives = 431/450 (95%), Gaps = 1/450 (0%)

- 35 Query: 3 EVSELRVQPQDLLAEQAVLGSIFISPEKLIMVREFISPDDFYKYSHKVIFRAMITLADRN 62
 EV+ELRVQPQDLLAEQ+VLGSIFISP+KLI VREFISPDDFYKY+HK+IFRAMITL+DRN
 Sbjct: 8 EVAELRVQPQDLLAEQSVLGSIFISPDKLIAREFISPDDFYKYAHKIIIFRAMITLSDRN 67
- 40 Query: 63 DAIDAATVRNILDQDGLQNLGGLGYIVELVNSVPTSANAEFYAKIVSEKAMLRDIISKL 122
 DAIDA T+R ILDDQ DLQ+IGGL YIVELVNSVPTSANAE+YAKIV+EKAMLRDII++L
 Sbjct: 68 DAIDATTIRTILDQDDLQSLGGLSYIVELVNSVPTSANAEYYAKIVAEEKAMLRDIIARL 127
- 45 Query: 123 TDTVMAY-EGNDSDEIIATAEKALVDINEHSNRSGFRKISDVLKVNENLELRSQQTSD 181
 T++VN+AY E +E+IA E+AL+++NEHSNRSGFRKISDVLKVNVE LE RS+QTS+
 Sbjct: 128 TESVNLAYDEILKPEEVIAGVERALIELNEHSNRSGFRKISDVLKVNVEALEARSKQTSN 187
- 50 Query: 182 VTGLPTGFRDLDRITTLGLHPDQLIILAARPAVGKTAFLVNLIAQNVGTRKQNRPAIFSLEM 241
 VTGLPTGFRDL+ITTLGLHPDQL+ILAARPAVGKTAFLVNLIAQNVGTRKQ + VAIFSLEM
 Sbjct: 188 VTGLPTGFRDLKITTGLHPDQVLILAARPAVGKTAFLVNLIAQNVGTRKQKKTVAIFSLEM 247
- 55 Query: 242 GAESLVDRMLAEGMVDHSLRTGQLTDQDWNNTTIAQGALADAPIYIDDTPGIKITEIR 301
 GAESLVDRMLAEGMVDHSLRTGQLTDQDWNNTTIAQGALA+APIYIDDTPGIKITEIR
 Sbjct: 248 GAESLVDRMLAEGMVDHSLRTGQLTDQDWNNTTIAQGALEAPIYIDDTPGIKITEIR 307
- Query: 302 ARSRKLSQEVDDGLGLIVIDYLQLISGTRPENRQQEVSEISRQLKILAKELKVPVIALSQ 361
 ARSRKLSQEVDD GLGLIVIDYLQLI+GT+PENRQQEVS+ISRQLKILAKELKVPVIALSQ
 Sbjct: 308 ARSRKLSQEVDDGGLGLIVIDYLQLITGT+PENRQQEVS+ISRQLKILAKELKVPVIALSQ 367
- Query: 362 LSRGVEQRQDKRPVLSDIRESGSIEQDADIVAFLYRDDYYRREGEEAEIIVEDNTVEVIL 421
 LSRGVEQRQDKRPVLSDIRESGSIEQDADIVAFLYRDDYYR+E ++AEE VEDNT+EVIL

-1161-

Sbjct: 368 LSRGVEQRQDKRPVLSDIRESGSIEQDADIVAFLYRDDYRKECDAAEEAVEDNTIEVIL 427

Query: 422 EKNRAGARGTVKLMFQKEYNKFSSIAQFEE 451
EKNRAGARGTVKLMFQKEYNKFSSIAQFEE

5 Sbjct: 428 EKNRAGARGTVKLMFQKEYNKFSSIAQFEE 457

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1039

- 10 A DNA sequence (GBSx1111) was identified in *S.agalactiae* <SEQ ID 3211> which encodes the amino acid sequence <SEQ ID 3212>. Analysis of this protein sequence reveals the following:

Possible site: 61
>>> Seems to have no N-terminal signal sequence

- 15 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.4909(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 20 The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3213> which encodes the amino acid sequence <SEQ ID 3214>. Analysis of this protein sequence reveals the following:

Possible site: 21
>>> Seems to have no N-terminal signal sequence

- 25 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.3467(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 30 An alignment of the GAS and GBS proteins is shown below.

Identities = 77/90 (85%), Positives = 84/90 (92%)

- 35 Query: 1 MSDAFADVAKMKKIKEDIKSHEGQMVLTLENGRKREKNKIGRLIEVYPSLFIVEYKDTA 60
MSDAF DVAKMKKIKEDI++HEGQ+VELTLENGRKREKNKIGRLIEVY SLFI+EY D++
Sbjct: 11 MSDAFTDVAKMKKIKEDIRAHEGQLVELTLENGRKREKNKIGRLIEVYSSLPFIEYSDSS 70

- Query: 61 AVPGAIDNTYVESYTYSDILTEKTLIRYFD 90
PGAIDN+YVESYTYSDILTEKTLIRY D
40 Sbjct: 71 DTPGAIDNSYVESYTYSDILTEKTLIRYLD 100

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1040

- 45 A DNA sequence (GBSx1112) was identified in *S.agalactiae* <SEQ ID 3215> which encodes the amino acid sequence <SEQ ID 3216>. This protein is predicted to be 30S ribosomal protein S4 (rpsD). Analysis of this protein sequence reveals the following:

Possible site: 27
>>> Seems to have no N-terminal signal sequence

- 50 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.2937(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

-1162-

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

5 >GP:AAC00397 GB:AF008220 ribosomal protein S4 [Bacillus subtilis]
Identities = 138/201 (68%), Positives = 158/201 (77%), Gaps = 1/201 (0%)

Query: 1 MSRYTGPSWKQSRRLGLSLTGTGKELARRNYVPGQHGPNNRSKLSEYGLQLAEKQKLRFS 60
M+RYTGPSWK SRRLG+SL+GTGKEL +R Y PG HGP R KLSEYGLQL EKQKLR
10 Sbjct: 1 MARYTGPSWKLSRRLGISLSGTGKELEKRPYAPGPHGPGQRKKLSEYGLQLQEKQKLRHM 60

Query: 61 YGLGEKQFRNLFVQATKAKEGTLGFNFMVLLERRLDNVVYRLGLATTRRQARQFVNHGHI 120
YG+ E+QFR LF +A K G G NFM+LL+ RLDNVVY+LGLA TRRQARQ VNHGHI
Sbjct: 61 YGVNERQFRTLFDKAGKLA-GKHGENFMILLDSRLDNVVYKLGRLATRRQARQLVNHGHI 119

15 Query: 121 LVDGKRVDIPSRYRVTGQVISVREKSMKVPAILLEAVEATLGRPAFVSFDAEKLEGSLTRL 180
LVDG RVDIPSY V PGQ I VREKS + I E+VE P +++FDAEKLEG+ TRL
Sbjct: 120 LVDGSRVDIPSYLKPGQTIGVREKSRNLSIIKESVEVNNFVPEYLTFDAEKLEGTFTRL 179

20 Query: 181 PERDEINPEINEALVVEFYNK 201
PER E+ PEINEAL+VEFY++
Sbjct: 180 PERSELAPEINEALIVEFYSR 200

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3217> which encodes the amino acid sequence <SEQ ID 3218>. Analysis of this protein sequence reveals the following:

25 Possible site: 27
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2937(Affirmative) < succ>
30 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 201/203 (99%), Positives = 201/203 (99%)

35 Query: 1 MSRYTGPSWKQSRRLGLSLTGTGKELARRNYVPGQHGPNNRSKLSEYGLQLAEKQKLRFS 60
MSRYTGPSWKQSRRLGLSLTGTGKELARRNYVPGQHGPNNRSKLSEYGLQLAEKQKLRFS
Sbjct: 1 MSRYTGPSWKQSRRLGLSLTGTGKELARRNYVPGQHGPNNRSKLSEYGLQLAEKQKLRFS 60

40 Query: 61 YGLGEKQFRNLFVQATKAKEGTLGFNFMVLLERRLDNVVYRLGLATTRRQARQFVNHGHI 120
YGLGEKQFRNLFVQATK KEGTLGFNFMVLLERRLDNVVYRLGLATTRRQARQFVNHGHI
Sbjct: 61 YGLGEKQFRNLFVQATKIKEGTLGFNFMVLLERRLDNVVYRLGLATTRRQARQFVNHGHI 120

45 Query: 121 LVDGKRVDIPSRYRVTGQVISVREKSMKVPAILLEAVEATLGRPAFVSFDAEKLEGSLTRL 180
LVDGKRVDIPSRYR PGQVISVREKSMKVPAILLEAVEATLGRPAFVSFDAEKLEGSLTRL
Sbjct: 121 LVDGKRVDIPSRYRVDPGQVISVREKSMKVPAILLEAVEATLGRPAFVSFDAEKLEGSLTRL 180

Query: 181 PERDEINPEINEALVVEFYNKML 203
PERDEINPEINEALVVEFYNKML
50 Sbjct: 181 PERDEINPEINEALVVEFYNKML 203

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1041

55 A DNA sequence (GBSx1113) was identified in *S.agalactiae* <SEQ ID 3219> which encodes the amino acid sequence <SEQ ID 3220>. Analysis of this protein sequence reveals the following:

Possible site: 29
>>> Seems to have no N-terminal signal sequence

-1163-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4067(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

5

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF98302 GB:AF243383 unknown; Orf3 [Lactococcus lactis subsp.
 lactis]

Identities = 46/97 (47%), Positives = 69/97 (70%)

10

Query: 1 MNLNDRLEKIEEMBEKYDSFKPRINALVEAIDDFQKHVEDYVKLREFYGSSEDWFRLESEQTE 60

M+ D I++ME KYD+F P + L++++ F Y +Y++LR FYGSE WF E +

Sbjct: 1 MDNKDIELIQQMENKYDTFMPVLTLNLDISVEKFNSIYNNYIELRNFYGSSEKWFYEMEIEK 60

15

Query: 61 NNLKCGVLSEDQLFDFIGEHNELVGQFLDMSSQMYRH 97

+KCGVL+EDQLFD I +HNEL+G LD++S+MY++

Sbjct: 61 IPVKCGVLTEQLFDMISDHNELGLVLLDLTSKMYKN 97

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3221> which encodes the amino acid
 sequence <SEQ ID 3222>. Analysis of this protein sequence reveals the following:

20

Possible site: 34

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

25

bacterial cytoplasm --- Certainty=0.3465(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

30

Identities = 48/98 (48%), Positives = 74/98 (74%)

Query: 1 MNLNDRLEKIEEMBEKYDSFKPRINALVEAIDDFQKHVEDYVKLREFYGSSEDWFRLESEQTE 60

M D+L +E+ME+ Y++F P++ L+EA+D F++HYE+Y LR FY S++WFRL+ Q

Sbjct: 1 MTKQDQLIVEKMEQTYEAFSPKLANLIEALDAFKEHYEYATLRNFYSSDEWFRLANQFW 60

35

Query: 61 NNLKCGVLSEDQLFDFIGEHNELVGQFLDMSSQMYRHL 98

+++ CGVLSED LFD IG+HN+L+ LD++ MY+H+

Sbjct: 61 DDIPCGVLSEDLFLDMIGDHNQLLADILD LAPIMYKHM 98

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 vaccines or diagnostics.

Example 1042

A DNA sequence (GBSx1114) was identified in *S.agalactiae* <SEQ ID 3223> which encodes the amino
 acid sequence <SEQ ID 3224>. Analysis of this protein sequence reveals the following:

45

Possible site: 50

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

50

bacterial cytoplasm --- Certainty=0.0965(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:EAB04438 GB:AP001509 transcriptional regulator (TetR/AcrR
 family) [Bacillus halodurans]

55

Identities = 47/181 (25%), Positives = 95/181 (51%), Gaps = 16/181 (8%)

Query: 4 DTRREKTKRAIEAAMITLLKQSFDEISTINLTKTAGISRSSFYTHYKDYEMIDQYQQS 63

-1164-

D R++ T+ ++ +++ L++++ I+ + A I+RS+FY+HY D Y+++ Q +
 Sbjct: 6 DRRKKYTRMLLKESLMKLMQEKPLSNITIKEICDLADINRSTFYSHYTDLYDLLYQIEDE 65
 Query: 64 LFNKV-EYIFDRNQFKKEDAL-----LEIFQFLDRESLFAALLTQNGTKEIQTYILNKLQ 117
 + + E + N K E+AL L ++ +RES L ++ G Q K
 Sbjct: 66 IIKDLSEALSSYNYTKDEEALQMTENLLVYIANNRESC-QTLFSEYGDPSFQ-----KKV 119
 Query: 118 LMLSKELPVVNP---DATKSDINRLYYSVYLSHAIFGVYQMWITRGKKESPQQITQVLLSL 175
 +ML+ + + P TK DI+ Y S+Y+ + + Q W+ G K+SP+++ ++++ L
 Sbjct: 120 MMLAHDHVIKTPLVGKHTKPDISE-YVSLYIVNGSIHIVQSWLKNGLKQSPKEMAELIILK 179

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3225> which encodes the amino acid sequence <SEQ ID 3226>. Analysis of this protein sequence reveals the following:

Possible site: 48
 >>> Seems to have an uncleavable N-term signal seq
 ----- Final Results -----
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:BAB04438 GB:AP001509 transcriptional regulator (TetR/AcrR
 family) [Bacillus halodurans]
 Identities = 47/180 (26%), Positives = 88/180 (48%), Gaps = 18/180 (10%)
 Query: 4 RKENTKQAILKAMVMLLKTESFDDITTVKLSKRAGISRSSFYTHYKDKYEMIDYYQQTFF 63
 RK+ T+ + ++++ L++ + +IT ++ A I+RS+FY+HY D Y+++ +
 Sbjct: 8 RKKYTRMLLKESLMKLMQEKPLSNITIKEICDLADINRSTFYSHYTDLYDLLYQIEDEII 67
 Query: 64 HKLEYIFEKKYQKQKQAFLEVFEEFL-----QREQLLSLLSANGTKEIQAFIINKVRLL- 117
 L K++ L++ E L + +L S G Q KV +L
 Sbjct: 68 KDLSEALSSYNYTKDEEALQMTENLLVYIANNRESCQTLFSEYGDPSFQ-----KKVMMLA 123
 Query: 118 ----ITDLQDKFSTEELSQTKEYQSIYLAHAFFGVCQSWIAKGGKESPQEMTQFVLKM 173
 I T L K + ++S EY S+Y+ + + QSW+ G K+SP+EM + ++K+
 Sbjct: 124 HDHVIKTPLVGKHTKPDISE----EYVSLYIVNGSIHIVQSWLKNGLKQSPKEMAELIILK 179

An alignment of the GAS and GBS proteins is shown below.

Identities = 100/179 (55%), Positives = 134/179 (73%), Gaps = 2/179 (1%)
 Query: 1 MVNDTRREKTKRAIEAAMITLLKDQSFDEISTINLTKTAGISRSSFYTHYKDKYEMIDQY 60
 MVN R+E TK+AI AM+ LLK +SFD+I+T+ L+K AGISRSSFYTHYKDKYEMID Y
 Sbjct: 1 MVN--RKENTKQAILKAMVMLLKTESFDDITTVKLSKRAGISRSSFYTHYKDKYEMIDYY 58
 Query: 61 QQSLEFNKVEYIFDRNQFKKEDALLEIFQFLDRESLFAALLTQNGTKEIQTYILNKLQML 120
 QQ+ F+K+EYIF++ KE A LE+F+FL RE L ++LL+ NGTKEIQ +I+NK++L++
 Sbjct: 59 QQTFFHKLEYIFEKKYQKQKQAFLEVFEEFLQREQLLSLLSANGTKEIQAFIINKVRLLI 118
 Query: 121 SKELPVVNPDATKSDINRLYYSVYLSHAIFGVYQMWITRGKKESPQQITQVLLSLLPQT 179
 + +L S + Y S+YL+HA FGV Q WI +GKKESPQ++TQ +L +L T
 Sbjct: 119 TTDLQDKFSTEELSQTKEYQSIYLAHAFFGVCQSWIAKGGKESPQEMTQFVLKMLTST 177

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1043

A DNA sequence (GBSx1115) was identified in *S.galactiae* <SEQ ID 3227> which encodes the amino acid sequence <SEQ ID 3228>. Analysis of this protein sequence reveals the following:

Possible site: 58

-1165-

>>> Seems to have no N-terminal signal sequence

5 INTEGRAL Likelihood = -10.35 Transmembrane 790 - 806 (787 - 808)
 INTEGRAL Likelihood = -7.32 Transmembrane 707 - 723 (703 - 725)
 INTEGRAL Likelihood = -7.11 Transmembrane 637 - 653 (630 - 659)
 INTEGRAL Likelihood = -6.32 Transmembrane 678 - 694 (672 - 698)
 INTEGRAL Likelihood = -1.44 Transmembrane 55 - 71 (55 - 73)
 INTEGRAL Likelihood = -0.22 Transmembrane 732 - 748 (730 - 748)

----- Final Results -----

10 bacterial membrane --- Certainty=0.5140(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

15 A related GBS nucleic acid sequence <SEQ ID 10287> which encodes amino acid sequence <SEQ ID 10288> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB12856 GB:Z99109 alternate gene name: yixE-similar to phage
 infection protein [Bacillus subtilis]
 Identities = 227/783 (28%), Positives = 387/783 (48%), Gaps = 60/783 (7%)
 20 Query: 45 KAIKSPKLWITMAGVALIPTLYNVIFLSSMWDPYGNTKNLPVAVVNQDKSAKLNKGTIS 104
 K I+ S KL I + + +P +Y+ +FL + WDPYG LPV VVNQDK A G+ +
 Sbjct: 9 KDIVTSKKLLIPIIIAILFVPLIYSGVFLKAYWDPYGTVDQLPVVVVNQDKGATYEGEKLQ 68
 25 Query: 105 IGDMDNLSKNDSLDFHFTT-AKRAKELEKGYHYMVITFPKDLRSKATTLMTKEPERL 163
 IG D+ L N++ D+HF+ ++ K+L YY+V+ P+D S+ A+T++ + P++L
 Sbjct: 69 IGDVLVKELKDNNNFDWHFSNDLDQSLKDLLNQKYLVVEIPEDFSKNASTVLDKNPKKL 128
 30 Query: 164 NITYKTTKGRSFSVASKMSETAANKLKDEVAESITGTYTESVFKNMGSMKTGINKAADGSQ 223
 ++ Y T G ++V + + E A +KLK V++ +T YT+ +F N + G++ A+ G++
 Sbjct: 129 DLKYHTNAGSNYVGATIGEKAIKDKASVSKEVTEQYTKVIFDNFKDIAKGLSDASSGAK 188
 35 Query: 224 ELLNGSNKLQDGSQTLTSLNLDVLASSSQTFSGGANKLNSGINLYTDGVGTLSNGLLETLS 283
 ++ +G+ ++GS L NL L S+ T S +L G T G+ +L + L D
 Sbjct: 189 KIDDGTDKAKNGSAQLKENLAKLKESTATISDKTAQLADGAAQVTSIGIQSLDSSLGKFQD 248
 40 Query: 284 GVTAYTTGVHKLSEGSQKLDDKSQALV-----EGSEKLTDLGLQQLSQATQLKPEQERT 336
 +L+ GS +L K L+ +G+ LT+GL QL+ Q E+
 Sbjct: 249 SSNQIYDKSSQLAAGSGBELTSKMNELLAGLQNVQKGPNTLNLGLDQLNSKVQEGSEKA 308
 45 Query: 337 LQNLSDG--LKNLQIITNLQSTATTDSDTNSKLFNFLSTIESSTKALMNTAAADKQKQM 394
 + + + L L + NL+ + T + +L +F +++++ +A N + +
 Sbjct: 309 AEKIINALDLTKLETAVNNLEKSETAMKEFKQLTDFENSLKNRDQAFKN--VINSSDFL 366
 50 Query: 395 TAVQST----SAFKSLTPEQQSQITSAVTGTPTSAE-TIAANISSNIENMKTVLSEASS 449
 TA Q + S K L ++ PT+ + A I S++E++K +++ +
 Sbjct: 367 TAEQKSQILNSVEKKLPQVDAPDFDQILSQLPTADQLPDIATIKSSLEDVKAQVAQVKAM 426
 55 Query: 450 APSN----NGSQNLQTLSGTANNVLKAIISDLDKIQKLPATKQLYQSSQTLTKGITDYT 505
 + NG++ +Q D I +L ++Y GSQ LT G T T
 Sbjct: 427 PEATSKLYNGAKTIQ-----DAIDRLTEGADKIYNGSQKLTGQTKLT 469
 Query: 506 NAVGQLRKGAIVTLDKSNQLISGTQKASQGAQTLDKSDQLRDGAGQLASGSDRIADGSN 565
 +G+ K + S QL++G S Q+ G +L GS ++ GS+
 55 Sbjct: 470 AGIGEYNKQFAKAKAGSEQLVTG-----SSQVSGGLFKLLDGSQKQVQSGSS 515
 Query: 566 KLAGGGHQLTDGLTELSGGVSQSSSLGKAGDQLSMVSVNKNANAVSSPVTIKHEDYDS 625
 KLA G L GL +L G +LSS L A DQ + + + PV K + S
 Sbjct: 516 KLAGGSASLDTGLGKLLDGTGELSSKLKDAADQTDGIDADDQTYGMFADPVKTKDDAIHS 575
 60 Query: 626 VDTNGVGMAPYMISVALMVVALSANVIFAKALSGKEPANRFSWAKNK--LLINGFIATL 682
 V G G+ PY++S+ L V + V+F + P N F W +K +++ G I +L
 Sbjct: 576 VPNYGTGLTPYILSMGLYVGGIMLTVVFPLEASGRPRNGFEWFFSKFNVMLVGLIIQSL 635
 65 Query: 683 -AATILFFAVQFIGLKPDPYKTYFIILLTAWTLMALVTALVGWDNRYGSFSLLLILLFQ 741

-1166-

AT+L IGL+ + + Y ++T+ +A++ L G F++++IL+ Q
 Sbjct: 636 IVATVLLIG--IGLEVESTWRFYVFTIITS LAFLAIQFLATTMGNPGRFIAVILVLQ 692
 Query: 742 LGSSAGTYPIELSPKFFQTIQPFPMYTSVSGLERETISLTGDVNHQWRMLVIFLVSSMIL 801
 5 LG+S GT+P+EL P F+Q I LPMTYS++G R IS GD + W+M + + ++++
 Sbjct: 693 LGASGTFPLELLPNFYQVIHGALPMTYSINGFRAVIS-NGDFGYMQMAGVLIGIALVM 751
 Query: 802 ALL 804
 L
 10 Sbjct: 752 IAL 754

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2017> which encodes the amino acid sequence <SEQ ID 2018>. Analysis of this protein sequence reveals the following:

Possible site: 26
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -9.29 Transmembrane 735 - 751 (729 - 754)
 INTEGRAL Likelihood = -5.79 Transmembrane 582 - 598 (580 - 601)
 INTEGRAL Likelihood = -3.66 Transmembrane 652 - 668 (650 - 669)
 20 INTEGRAL Likelihood = -2.97 Transmembrane 14 - 30 (14 - 34)
 INTEGRAL Likelihood = -2.66 Transmembrane 623 - 639 (622 - 641)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.4715(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 25 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 360/779 (46%), Positives = 508/779 (64%), Gaps = 32/779 (4%)

30 Query: 40 MLDELKAIKSPKLWITMAGVALIPTLYNVIFLSSMWDPYGNTKNLPVAVVNQDKSAKLN 99
 ML+ELK +IK+PKL ITM GVAL+P LYN+ FL SMWDPYG +LP+AVVN DK AK
 Sbjct: 1 MLEELKTLIKNPKLMITMIGVALVPALYNLSFLGSMWDPYGRVNDLPVAVVNHDKPAKRA 60
 Query: 100 GKTISIGKDMEDNLSKNDSDLFHFTAKRAEKELEKGHYVMVITFPKDL SRKATTLMTEK 159
 35 K+++IG DM D +SK+ L++HF +AK+A++ L++G YMVIT P+DLS++A TL+ +
 Sbjct: 61 DKSLTIGNDMVDKMSKSKDLEYHFVSAKQAEGLKEGDYVMVITLPEDLSQRAATLLNPE 120
 Query: 160 PERLNTITYKTTKGRSFVASKMSETAANKLKDEVAESITGTYTESVFKNMGSMKTGINKAA 219
 P++L I Y+T+KG VA+KM ETA KLK+ V+++IT TYT +VF +M +++G+ +A+
 40 Sbjct: 121 PQKLITIRYQTSKGHGMAAKMGETAMAKLKESVSQNTKTYTSAVFSSMTDLQSGLEAS 180
 Query: 220 DGSQELLNGSNKLQDGSQTLTSLNLDVLASSQTFSGGANKLNSGINLYTDGVTLSNGLE 279
 GSQ L +G+ Q GSQTL++NL L +SQ F G +L SG+ YTDGV + NGL
 45 Sbjct: 181 AGSQALASGAKTAQAGSQTLSTNLAAITGASQQFQQGTGRLTSTYTDGVNQVKNGLG 240
 Query: 280 TLDGVTAYTTGVHKLSEGSQKLDKQALVEGSEKLTDLGLQQLSQATQLKPEQERTLQN 339
 TLS + Y GV +LS+G+ +L+ GL QL+QAT L E+ + +Q+
 Sbjct: 241 TLSTDIPNYLNGVSRLSQGASQLNQ-----GLSQLTQATTLSDKAKGIQS 286
 Query: 340 LSDGLKNLNQIITNLQSTATTDSDTN---SKLFNFLTIESSTKALMNTAAADKQKQMTA 396
 L GL LNQ I L + +T N +L N L I + K ++ A + ++++A
 50 Sbjct: 287 LIVGLEPVLNQGIQQLNTELSTLQPPNLDLGNLGAIAQAQVIAEETAQAQNEELSA 346
 Query: 397 VQSTSAFKSLTPEQSQITSAVTGTPTSATETIAAN-ISSNIENMKTVLSEASSAPSNNNG 455
 +Q+TS ++SLT EQQ ++ +A++ + S AA I S+++ + T L S S
 55 Sbjct: 347 LQATSVYQSLTAEQQGELAAALSQSDKQSTVSAQTILSSVQTLSTSLQSLSQEDQSKQL 406
 Query: 456 SONLQTLSTGTANNLVLKAIKSLDLDKIQLPTATKQLYQGSQTLTKGITDYTNV----GQL 511
 Q + ++ AN Q LP A+ L + S L K V QL
 60 Sbjct: 407 EQLKAEVAQIANQ-----SNQALPGASSALTELSTGLAKVNGSLNQVLPQSNQL 456
 Query: 512 RKGAVTLDKSNQLISGTQKASQGAQTLDSKSDQLRDGAGQLASGSDRIADGSNKLGGG 571
 G L+ + + SG K S+GA L SKS +L DG+ QL+ G+ ++ADGS++L+ GG
 65 Sbjct: 457 TTGLAQLNRYNTAIGSGVIKLSEGANALSSKSGELLGSHQLSEGATKLDGSSQLSQGG 516

-1167-

Query: 572 HQLTDGLTELSGGVSQLSSSLGKAGDQLSMVSVNKNANAVSSPVTIKHEDYDSVDITNGV 631
 HQLT GLTELS G+S L+ SL KA QLS+VSV NA AV+ P+ + +D D V TNG+
 Sbjct: 517 HQLTSGLTSTGLSTLNGSLAKASQQLSLVSVDKNAKAVAKPLVLNEKDKDGVKTNGI 576

5 Query: 632 GMAPYMISVALMVVALSANVIFAKALSGKEPANRFSWAKNKLINGFTIATLAATILFFAV 691
 GMAPYMI+V+LMVVALS NVIFA +LSG+ +++ WAK K +INGFI+T+ + +L+ A+
 Sbjct: 577 GMAPYMIASVSLMVVALSTNVIFANSLSGRPVKDKWDWAKQKFVINGFISTMGSIIVLYLAI 636

10 Query: 692 QFIGLKPDPGKTYFIILLTAWTLMALVTALVGWDNRYGSFSLLLILLFQLGSSAGTYPI 751
 Q +G + Y +T I+L+ WT MALVTALVGWD+RYGSF SL++LL Q+GSS G+YPI
 Sbjct: 637 QLLGFEARYGMETLGFIMLSGWTFMALVTALVGWDDRYGSFASLVMLLLQVGSSGGSYPI 696

Query: 752 ELSPKFFQTIQPFPLMTYVSGLRETISLTGDNHQRMLVIFLVSSMILALIYRKQE 810
 ELS FFQ + PFLPMY VSGLR+TISL+G + + ++L FL++ M+LALLIYR ++
 15 Sbjct: 697 ELGAFFQKLHPFLPMYVVSGLRQTISLSGHIGVEVKVLTGFLAFMVLALIYRPPK 755

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1044

20 A DNA sequence (GBSx1116) was identified in *S.agalactiae* <SEQ ID 3229> which encodes the amino acid sequence <SEQ ID 3230>. Analysis of this protein sequence reveals the following:

Possible site: 35
 >>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2664(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

30 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1045

35 A DNA sequence (GBSx1117) was identified in *S.agalactiae* <SEQ ID 3231> which encodes the amino acid sequence <SEQ ID 3232>. Analysis of this protein sequence reveals the following:

Possible site: 60
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -9.45 Transmembrane 48 - 64 (45 - 69)
 40 INTEGRAL Likelihood = -1.49 Transmembrane 71 - 87 (71 - 87)

----- Final Results -----
 bacterial membrane --- Certainty=0.4779(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 45 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9441> which encodes amino acid sequence <SEQ ID 9442> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

50 >GP:AAA25222 GB:M87483 ORF 1 [Lactococcus lactis]
 Identities = 50/88 (56%), Positives = 66/88 (74%), Gaps = 1/88 (1%)
 Query: 2 TGIKIFMSKEELSYLPVLIKLFKNQGVYNGLIGLFLLVGLYISQNG-EIVAVFLINVLVA 60

-1168-

T ++F+M KEEL V LFKNQG+YNGLIGL L+Y ++ S Q EIV + LI ++LVA
 Sbjct: 32 TSRVFNMGKEELERSVQTLFKNQGIYNGLIGLGLIYAIFFSSAQLEIVRLLLIYIILVA 91

Query: 61 IYGALTVDKKILLKQGGLPILALLTFLF 88

5 +YG+LT +KKI+L QGGL ILAL++ F

Sbjct: 92 LYGSILTSNKKIILTQGGILAILALISSFF 119

No corresponding DNA sequence was identified in *S.pyogenes*.

10 A related GBS gene <SEQ ID 8719> and protein <SEQ ID 8720> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 8

McG: Discrim Score: 4.19

GvH: Signal Score (-7.5): -3.99

Possible site: 38

15 >>> Seems to have an uncleavable N-term signal seq

ALOM program count: 3 value: -9.45 threshold: 0.0

INTEGRAL Likelihood = -9.45 Transmembrane 87 - 103 (84 - 108)

INTEGRAL Likelihood = -1.49 Transmembrane 110 - 126 (110 - 126)

20 INTEGRAL Likelihood = -0.37 Transmembrane 13 - 29 (13 - 29)

PERIPHERAL Likelihood = 0.47 65

modified ALOM score: 2.39

*** Reasoning Step: 3

25 ----- Final Results -----

bacterial membrane --- Certainty=0.4779(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

30 The protein has homology with the following sequences in the databases:

ORF00610(328 - 681 of 981)

SP|Q02009|YTRP_LACLA(1 - 119 of 119) HYPOTHETICAL 13.3 KDA PROTEIN IN TRPE 5'REGION.

GP|551879|gb|AAA25222.1|M87483 ORF 1 {Lactococcus lactis} PIR|S35123|S35123 hypothetical protein (trpE 5' region) - Lactococcus lactis subsp. lactis

35 %Match = 19.9

%Identity = 58.8 %Similarity = 77.3

Matches = 70 Mismatches = 26 Conservative Sub.s = 22

114 144 174 204 234 264 294 324
 40 SPKFFQTIQPFLEPMTYSVSGLRITSLTGDVNHQWRMLVIFLVSSMILALLIYRKQED**KVSSDRLTV*YGMISKYLGG

354 384 414 444 474 504 534 561

DMSTLTIIITATLTALEHFYIMYLETLATQSNMTGKIFMSKEELSYPVVKLFKNQGVYNGLIGLFLYGLYISQNG-EI

45 | : ||| : : | ||| ||||| : || : || : ||| : ||||| : ||||| : || : || : |||

MTILTIILSLVALEFFYIMYLETFATSSKTTSRVFNMGKEELERSVQTLFKNQGIYNGLIGLGLIYAIFFSSAQLEI

10 20 30 40 50 60 70

591 621 651 681 711 741 771 801

VAVFLINVLVAIYGALTVDKKILLKQGGLPILALLTFLF*YYLAVRFS*TAFSNHFFLIQV*VICL*K*YNITTNSK

50 | : || : || : || : || : || : || : || : || : || : || : || : || : || : || : ||

VRLLLIYIILVALYGSILTSNKKIILTQGGILAILALISSFF

90 100 110

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1046

A DNA sequence (GBSx1118) was identified in *S.agalactiae* <SEQ ID 3233> which encodes the amino acid sequence <SEQ ID 3234>. Analysis of this protein sequence reveals the following:

Possible site: 41

-1169-

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.3140(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10285> which encodes amino acid sequence <SEQ ID 10286> was also identified.

10 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB12447 GB:Z99107 similar to arylesterase [Bacillus subtilis]
 Identities = 37/91 (40%), Positives = 56/91 (60%)

15 Query: 13 KDGSDIYYRVVGQGQPIVFLHGNLSRRYFDKQIAYFSKYYQVIVMDSRGHGKSHAKLNT 72
 +D + +YY G G PI+F+HG +S ++F KQ + S YQ I +D RGHG+S L+
 Sbjct: 7 EDQTRLIYYETHGSGTPILFIHGVLMSGQFFHKQFSVLSANYQCIRLDLRGHGESDKVLHG 66

Query: 73 ISFRQIAVDLKDILVHLEIDKVLVGHSDGA 103
 + Q A D+++ L +E+D V+L G S GA
 20 Sbjct: 67 HTISQYARDIREFLNAMELDHVVLGWSMGA 97

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

25 Example 1047

A DNA sequence (GBSx1119) was identified in *S.agalactiae* <SEQ ID 3235> which encodes the amino acid sequence <SEQ ID 3236>. This protein is predicted to be an integral membrane protein. Analysis of this protein sequence reveals the following:

Possible site: 58

30 >>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -12.90	Transmembrane	14 - 30 (9 - 41)
INTEGRAL	Likelihood = -9.71	Transmembrane	451 - 467 (447 - 472)
INTEGRAL	Likelihood = -9.18	Transmembrane	234 - 250 (229 - 257)
INTEGRAL	Likelihood = -8.07	Transmembrane	56 - 72 (46 - 77)
35 INTEGRAL	Likelihood = -8.01	Transmembrane	490 - 506 (484 - 512)
INTEGRAL	Likelihood = -5.84	Transmembrane	414 - 430 (412 - 436)
INTEGRAL	Likelihood = -4.99	Transmembrane	136 - 152 (135 - 159)
INTEGRAL	Likelihood = -4.14	Transmembrane	213 - 229 (211 - 232)
INTEGRAL	Likelihood = -4.14	Transmembrane	365 - 381 (364 - 382)
40 INTEGRAL	Likelihood = -2.66	Transmembrane	393 - 409 (391 - 412)
INTEGRAL	Likelihood = -1.06	Transmembrane	168 - 184 (167 - 184)
INTEGRAL	Likelihood = -0.64	Transmembrane	275 - 291 (275 - 291)
INTEGRAL	Likelihood = -0.32	Transmembrane	328 - 344 (328 - 345)
45 INTEGRAL	Likelihood = -0.27	Transmembrane	821 - 837 (821 - 837)

----- Final Results -----

50 bacterial membrane --- Certainty=0.6158(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10283> which encodes amino acid sequence <SEQ ID 10284> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

55 >GP:BAA24464 GB:D85082 Yfix [Bacillus subtilis]
 Identities = 190/596 (31%), Positives = 324/596 (53%), Gaps = 31/596 (5%)

-1170-

Query: 246 IVSLIPGGLGSFELVLTGFAAEGLPKETVVAWLLLYRLAYYIIPFFAGIYFFIHYLGSQ 305
 ++SL+PGG GSF+L+ G G +E +V ++LYRLAY IPF G++F L
 Sbjct: 1 MISLVPGGFGSFDLLFLGMEQLGYHQEAIVTSIVLYRLAYSFIPFILGLFFAAGDLTEN 60

Query: 306 INQRYENVPK-----ELVSTVLQTMVSHLMRILG---AFLIFSTAFFENITYIMWLQKLG 357
 +R E P+ E + +L + L+RIL + ++F + + + +L
 Sbjct: 61 TMKRLNPNRIAPAIETTNVLLVVQRAVLVRILQGSLSLIVFVAGLIVLASVSLPIDRLT 120

Query: 358 LDP-LQEQMLWQFFGGLLGVCIFILLARTID--QKVKNAPPIAIWITLTLFYLNLGHISW 414
 + P + L F GL L ILL I+ ++ K ++ +AI + + L ++
 Sbjct: 121 VIPHIPRALLLFLNGLSLSSALILLILPIELYKRTKRSYTMAITALVGGFVFSFLKGLNI 180

Query: 415 RLSFWFILLLLGLLVIKPTLYKKQFIYSWEERIKDGIIVSLMGVLFY----IAGLLFP 470
 F ++++ L+++K ++Q Y+ + I V+L V + IAG ++
 Sbjct: 181 SAIFVLPMIIVLVLLKKQFVREQASYTLGQLI----FAVALFTVALFNYNLIAGFIWDR 236

Query: 471 RAHITGGSIERLHYIIAWEPIALATL---ILTLVYLCLVKILQKSCQIGDVFNVDRYK 526
 + + +++ + I AT+ I+ L +L + ++ IG+ + +R
 Sbjct: 237 MKKV---LRHEYFVHSTSHITHTATIMAIIVPLFFLIPTVVYHKRTKPIGEKADPERLA 292

Query: 527 KLLQAYGGSSDSGLAFINDKRLYWYQKNGEDCVAFQFVIVNNKCLIMGEFAGDDTYIREA 586
 L GG++ S L FL DKR Y + +G + F + + +++G+P+G
 Sbjct: 293 AFLNEKGGNALSHLGLGDKRFY-FSSDGNALLLFGKIA--RRLVVLGDPGSGQRESFPLV 349

Query: 587 IESFIDDADKLDYDLVYFISIGQKLTLLLEHYGFDFMKVGEDALVNLETFTLKGNKYKPF 646
 +E F+++A + + ++FY I ++ L H++G++F K+GE+A V+L TFL G K R
 Sbjct: 350 LEEFLNEAHQKGFVLFYQIEREDMALYHDFGYNFFKLGEAYVDLNTFTLTGKKKAGLR 409

Query: 647 NALNRVEKDGIFYEVVQSPHSQELLNSLEEISNTWLEGRPEKGFSLGYFNKDYFQQAPIA 706
 NR E++ + F V P S L L++IS+ WL + EKGFSLG+F+ Y Q+APIA
 Sbjct: 410 AINNRFEREETTFHVDHPPFSDAFLEELKQISDEWLGSKEKGFSLGFPDPSYLQKAPIA 469

Query: 707 LVKNAEHEVVAFANIMPNYEKSIISIDLMRHDKQKIPNGVMDFLFLSLFSYYQEKGYHYF 766
 +KNAE E+VAFAN+MP Y++ IS+DLMR+ + PNG+MD LF+ +F + +E+G F
 Sbjct: 470 YMKNAEGEIVAFANVMPYQEGEISVDLMRY-RGDAPNGIMDALFIRMFLWAKEEGCTSF 528

Query: 767 DLGMAPLSGVGRVETSFAKERMAYLVYHFGSHFYFNLGHYKKKFTPLWSERYIS 822
 ++GMAPL+ VG TSF ER A ++++ + YSF+GL +K+K+ P W +Y++
 Sbjct: 529 NMGMAPLANVGTAFTSFWSERFAAVIFNNVRYMYSFSGLRAFKEKYKPEWRGKYLA 584

No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8721> and protein <SEQ ID 8722> were also identified. Analysis of this protein sequence reveals the following:

45 Lipop: Possible site: -1 Crend: 9
 McG: Discrim Score: 9.22
 GvH: Signal Score (-7.5): -7.66
 Possible site: 58

>>> Seems to have an uncleavable N-term signal seq

50 ALOM program count: 14 value: -12.90 threshold: 0.0

INTEGRAL	Likelihood = -12.90	Transmembrane	14 - 30 (9 - 41)
INTEGRAL	Likelihood = -9.71	Transmembrane	451 - 467 (447 - 472)
INTEGRAL	Likelihood = -9.18	Transmembrane	234 - 250 (229 - 257)
INTEGRAL	Likelihood = -8.07	Transmembrane	56 - 72 (46 - 77)
55 INTEGRAL	Likelihood = -8.01	Transmembrane	490 - 506 (484 - 512)
INTEGRAL	Likelihood = -5.84	Transmembrane	414 - 430 (412 - 436)
INTEGRAL	Likelihood = -4.99	Transmembrane	136 - 152 (135 - 159)
INTEGRAL	Likelihood = -4.14	Transmembrane	213 - 229 (211 - 232)
INTEGRAL	Likelihood = -4.14	Transmembrane	365 - 381 (364 - 382)
60 INTEGRAL	Likelihood = -2.66	Transmembrane	393 - 409 (391 - 412)
INTEGRAL	Likelihood = -1.06	Transmembrane	168 - 184 (167 - 184)
INTEGRAL	Likelihood = -0.64	Transmembrane	275 - 291 (275 - 291)
INTEGRAL	Likelihood = -0.32	Transmembrane	328 - 344 (328 - 345)
INTEGRAL	Likelihood = -0.27	Transmembrane	821 - 837 (821 - 837)
65 PERIPHERAL	Likelihood = 1.06		558

-1172-

AVIFNNVRYMYSFSGLRAFKEKYKPEWRGKYLAYRKNRSLSVTMFLVTRLIGKSKKDSV
 610 620 630 640 650

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1048

A DNA sequence (GBSx1120) was identified in *S. agalactiae* <SEQ ID 3237> which encodes the amino acid sequence <SEQ ID 3238>. This protein is predicted to be choline transporter. Analysis of this protein sequence reveals the following:

10 Possible site: 37
 >>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -10.24	Transmembrane	28 - 44 (22 - 47)
INTEGRAL	Likelihood = -8.81	Transmembrane	178 - 194 (176 - 204)
INTEGRAL	Likelihood = -7.22	Transmembrane	81 - 97 (63 - 105)
15 INTEGRAL	Likelihood = -3.50	Transmembrane	209 - 225 (206 - 226)
INTEGRAL	Likelihood = -3.13	Transmembrane	64 - 80 (63 - 80)
INTEGRAL	Likelihood = -2.44	Transmembrane	156 - 172 (153 - 172)
INTEGRAL	Likelihood = -0.64	Transmembrane	137 - 153 (137 - 153)

20 ----- Final Results -----
 bacterial membrane --- Certainty=0.5097(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

25 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAD45530 GB:AF162656 choline transporter [Streptococcus pneumoniae]
 Identities = 326/505 (64%), Positives = 409/505 (80%), Gaps = 1/505 (0%)

30 Query: 1 MTTLTITTFQERFGDWTQSLIEHLQLSLLTLILATLIAIPLGIIISHYKKISHVVLQITGI 60
 MT LI TFQ+RF DW +L +HLQLSLLTL+LA L+AIPL + + +++K++ VLQI GI
 Sbjct: 1 MTNLIATFQDRFSDWLTALSQHLQLSLLTLLLAILLAIPLAVFLRYHEKLADWVLQIAGI 60

35 Query: 61 FQTIPSLALLGLFIPFPMGIGTVPVAVALIYALFPILQNTVTVLMQIDANLIEAATAFGM 120
 FQTIPSLALLGLFIP MGIGT+PA+ AL+IYA+FPILQNT+T L ID NL EA AFGM
 Sbjct: 61 FQTIPSLALLGLFIPLMGIGTILPALVALVIYALFPILQNTITGLKGIDPNLQEAGIAFGM 120

40 Query: 121 TRWERLKKFELALSMPVLIISGIRTASVMIIGTATLASLIGAGGLGSFILLGIDRNNPSLI 180
 TRWERLKKFE+ L+MPVI+SGIRTA+V+IIGTATLA+LIGAGGLGSFILLGIDRNN SLI
 Sbjct: 121 TRWERLKKFEIPLAMPVIMSGIRTAAVLIIGTATLAALIGAGGLGSFILLGIDRNNASLI 180

45 Query: 181 LIGAISSAVLAIIFSGILGLEKARLRTIAVSGILLLAGLGLSYAPKWMPGTNTATITVA 240
 LIGA+SSAVLAI F+ L+ ++EKA+LRTI L+ LGLSY+P + + +A
 Sbjct: 181 LIGALSSAVLATAFNFLLKVMKAKLRTIFSGFALVALLLGLSYSPALLVQKEKENLVIA 240

50 Query: 241 GKLGTEPDILINMYKELIEDQTDIKVKLKNPFGKTTFLYQALKSGDIDLYPEFTGTITSS 300
 GK+G EP+IL NMYK LIE+ T + +KPNFGKT+FLY+ALK GDID+YPEFTGT+T S
 Sbjct: 241 GKIGPEPEILANMYKLLIEENTSMATVKNPFGKTSFLYEALKKGDIDDIYPEFTGTVTS 300

55 Query: 301 LLKNPPKVSNNPKQVYNLAKNGILKQDKLSLLSPMAYQNTYAVAVKKDYAEANQLKNISD 360
 LL+ PKVS+ P+QVY +A++GI KQD L+ L PM+YQNTYAVAV K A+ LK ISD
 Sbjct: 301 LLQPSPKVSHEPEQVYQVARDGIAKQDHLAYLKPMQSYQNTYAVAVPKKIAQEYGLKTISD 360

60 Query: 361 LKKLD-KLKAGFTLEFKDREDGSIGLQKHYNLNDISTLEPALRYQAINSKDVNIIDAYS 419
 LKK++ +LKAGFTLEF DREDG+ GLQ YGLNL+++T+EPALRYQAI S D+ I DAYS
 Sbjct: 361 LKKVEGQLKAGFTLEFNDREDGNKGLQSMYGLNLNVATIEPALRYQAIQSGDIQITDAYS 420

Query: 420 TDESLIQYQLQILKDDKHLFPPYQGAPLLRQDTIKKYPQVKKALNKLAGHITEKEMQEMN 479
 TD+EL +Y LQ+L+DDK LFPYQAGAPL+++ +KK+P++++ LN LAG ITE +M ++N
 Sbjct: 421 TDAELERYDLQVLEDDKQLFPPYQAGAPLMKEALKKHPELERVLNLTLAGKITESQMSQLN 480

Query: 480 YQVAVKHKSAATVAKQYLKAHHIHK 504

-1173-

YQV V+ KSA VAK++L+ ++K
 Sbjct: 481 YQVGVEGKSAKQVAKEFLQEQLLK 505

There is also homology to SEQ ID 636.

- 5 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1049

- 10 A DNA sequence (GBSx1121) was identified in *S.agalactiae* <SEQ ID 3239> which encodes the amino acid sequence <SEQ ID 3240>. This protein is predicted to be choline transporter (opuBA). Analysis of this protein sequence reveals the following:

Possible site: 59
 >>> Seems to have no N-terminal signal sequence

- 15 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2345(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

- 20 >GP:AAD45529 GB:AF162655 choline transporter [Streptococcus pneumoniae]
 Identities = 139/236 (58%), Positives = 178/236 (74%)
 Query: 1 MISFENVSKSYGDHTIIDNISCHIQRGEFFVLVGASGSGKTTILKMINRLIEPSQGAIL 60
 MI ++NV+ Y + ++ +++ I+ GEF VLVG SGSGKTT+LKMINRL+EP+ G I +
 25 Sbjct: 1 MIEYKNVALRYTEKDVLRDVLNQLIEDGEFMVLVGPSGSGKTTMLKMINRLLEPTDGNIIYM 60
 Query: 61 DGENITSLDLRQLRLETGYVLQIALLFPNLTVGNIELIPEMKGWSKGDQKKAASDLLDK 120
 DG+ I D R+LRL TGYVLQ IALFPNLTV ENI LIPEMKGWSK + K +LL K
 30 Sbjct: 61 DGKRIKDYDERELRLSTGYVLQIALLFPNLTVAENIALIPEMKGWSKEITKKTEELLAK 120
 Query: 121 VGLPAKDYFNRYPHELSGGEQQRIGILRAIVAKPKVLLMDEPFSSALDPISRRQLQDITKQ 180
 VGLP +Y +R P ELSGGEQQR+GI+RA++ +PK+ LMDEPFSSALD ISR+QLQ +TK+
 35 Sbjct: 121 VGLPVAEYGHRLPSELSGGEQQRVGIVRAMIGQPKIFLMDEPFSSALDAISRKQLQVLTKE 180
 Query: 181 LQSELGITLVFVTHDMKEAMRLADRICVIKEGKIVQLDRPEIIQNPNPSDQFVRTLF 236
 L E G+T +FVTHD EA++LADRI V+++G+I Q+ PE I P+ FV LF
 Sbjct: 181 LHKEFGMTTIFVTHDTDEALKLADRIAVLQDGEIRQVANPETILKAPATDFVADLF 236

There is also homology to SEQ ID 644.

- 40 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1050

- 45 A DNA sequence (GBSx1122) was identified in *S.agalactiae* <SEQ ID 3241> which encodes the amino acid sequence <SEQ ID 3242>. This protein is predicted to be two-component response regulator. Analysis of this protein sequence reveals the following:

Possible site: 61
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -5.52 Transmembrane 49 - 65 (46 - 66)

- 50 ----- Final Results -----
 bacterial membrane --- Certainty=0.3208(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

-1174-

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB06434 GB:AP001516 two-component response regulator [Bacillus halodurans]
Identities = 101/305 (33%), Positives = 152/305 (49%), Gaps = 31/305 (10%)

5 Query: 1 MKFYIIDDDPTITMILQDIIE-EDFNNTVVRVNVSSKAYNELLIADVDIVLIDLLMPIL 59
M F+I DDD T+ IL IIE E V + S L I VDI+LIDLLMP
Sbjct: 1 MNFFITDDDVTVRSILAQIIEDEQLGQVVGEAEDGSELDGKRLNIKQVDILLIDLLMPNC 60

10 Query: 60 DGVTLVQKIYKQSDLKFMISQVKDNDLRQEAYKAGIEFFINKPINIIEVKSVVKRVTD 119
DG+ +QKI K K IMISQ++ +L EAY GIE +I KPIN IEV SV+++V +
Sbjct: 61 DGLEAIQKI-KPEFKGKIIMISQIESKELISEAYLLGIEHYIMKPINKIEVLSVIRKVIN 119

15 Query: 120 TIEMQKKLNTIQNLLENTPSYQKPITTSNLT---KIRS---ILSYLGITSETAYTDIL 171
+++ L IQ L N P ++ I+S +LS LGI E+ D++
Sbjct: 120 HTRLBQSLYDIQKSLSNVLQSGIPTQVNDQVHFHDSIKSYGQYLLSELGIAGESGSKDLM 179

20 Query: 172 NICELLLKQELNF-----AQFDFQKELSIDE-----HQQKIILQRIIRRAVKK 213
NI L E + A D ++L+ ++ + K QR+RRAV +
Sbjct: 180 NILMFLYTYEKEYSFEKGFPALKDIFEQLASEKLGAADERDVRREVKAQKQVRRAVYQ 239

25 Query: 214 AMINMAHLYIDDFENELTLQYANALFGFQNIHNEAQLIQGK---SMYGGKISLKHFFDEL 270
++ ++A L + DF N +YA+ F F + ++ ++ + S +I++K F L
Sbjct: 240 SLEHVASLGLIDFSNPKFEEYASHFFDFS VVRSKMTELKNETSSSYTSARINVKKFTQAL 299

Query: 271 ILQSK 275
++K
Sbjct: 300 YYEAK 304

30 There is homology to SEQ ID 460.

A related GBS gene <SEQ ID 8723> and protein <SEQ ID 8724> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 8
McG: Discrim Score: -7.05
35 GvH: Signal Score (-7.5): -6.58
Possible site: 61
>>> Seems to have no N-terminal signal sequence
ALOM program count: 1 value: -5.52 threshold: 0.0
40 INTEGRAL Likelihood = -5.52 Transmembrane 49 - 65 (46 - 66)
PERIPHERAL Likelihood = 7.37 155
modified ALOM score: 1.60
*** Reasoning Step: 3
45 ----- Final Results -----
bacterial membrane --- Certainty=0.3208(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

50 The protein has homology with the following sequences in the databases:

ORF00604(307 - 1125 of 1431)
EGAD[137180|146289(3 - 304 of 310) hypothetical protein {Bacillus cereus}
GP|1769946|emb|CAA67094.1||X98455 orf1 {Bacillus cereus}
%Match = 12.7
55 %Identity = 34.1 %Similarity = 53.0
Matches = 95 Mismatches = 123 Conservative Sub.s = 53

168 198 228 258 288 318 348 375
*C*W*YLSRNRAIPRAYFNRAISRNDNCLS*SAKWNNTYVIP*KSI*VRR*YVKFYIIDDDPTITMILQDIIEE-DFN
60 :||:| :| |||:| :
MFYIIVDDDEVERSMLSQIIEDGLG
10 20

-1175-

```

405      435      465      495      525      555      585      615
NTVVRVNVNVS KAYNELLIADVDIVLIDLLMPILDGVTLVQKIYQKRS DLKFIMISQVKDNDLRQEAYKAGIEFFINKPI
:   :   :   :   |||::|||::|||: |:: |:: |   |   |||::|   |   ||| |:::| ||:
5 EVIGES EDGAFVEAEQLNKKVDILFIDLLMPMRDGIETVRHI-ASSFTGKIIMISQVESKQLIGEAYTLGVEYYITKPL
      40      50      60      70      80      90      100

645      675      705      753      771      801      831
NIIIEVKS VVKRVTDTIEMQKGLNTIQLLENTPSYQKP---ITTSNLTKI---RSILSYLGITSETAYTDILNICELL
|   |   |   |::| : | :: : | | | : | | |   |   |   | : | : | | |   | : | : | |
10 NKIEVVSVVRK VIERIRLERSIYDIQKSLNNVFQWEKPMRSETVQEEKISDSGRFLLAELGIAGENGSKDLLSMLEYL
      120     130     140     150     160     170     180

861      894      924      954      984      1014
LKQELNFAQFDFQKELSID-----EHQQKIILQRIIRAVKKAMINMAHLYIDDFENELTLQYANAL
| |   | : |   |   | : |   | : | : | : : : | : | |   | | :
15 YGQE-KAQTFEFGFPALKDIFHQITLKKLGEIASDADIEKEKKASEQVRRAIYQSLNHLASLGLTDFSNPKFESYAPKF
      200     210     220     230     240     250     260

1071     1095     1125     1155     1185     1215     1245
PGFQNIHNE-AQLIQGKSMYGGKISL--KHFFDELILQSKTF*DLFKHGLIYVNHPKTFLFINLQQTPLCPQGVCF*F
| | :   : : : | : | | | : : |
20 P DFTVVRKRMTEMTKDG VATSGHIRINTKKFIQVLYFEAKRLMEIE
      280     290     300     310

```

SEQ ID 8724 (GBS356) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell
 25 extract is shown in Figure 73 (lane 3; MW 34kDa). It was also expressed in *E.coli* as a GST-fusion
 product. SDS-PAGE analysis of total cell extract is shown in Figure 81 (lane 8; MW 59kDa).

GBS356-GST was purified as shown in Figure 216, lane 7.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 vaccines or diagnostics.

30 Example 1051

A DNA sequence (GBSx1123) was identified in *S.agalactiae* <SEQ ID 3243> which encodes the amino
 acid sequence <SEQ ID 3244>. Analysis of this protein sequence reveals the following:

```

Possible site: 26
>>> Seems to have a cleavable N-term signal seq.
35  INTEGRAL    Likelihood = -6.48    Transmembrane  149 - 165 ( 147 - 172)
    INTEGRAL    Likelihood = -5.20    Transmembrane   37 - 53 ( 29 - 55)
    INTEGRAL    Likelihood = -2.50    Transmembrane  126 - 142 ( 126 - 142)
    INTEGRAL    Likelihood = -2.13    Transmembrane   62 - 78 ( 60 - 78)
    INTEGRAL    Likelihood = -0.64    Transmembrane  314 - 330 ( 314 - 330)
40  INTEGRAL    Likelihood = -0.11    Transmembrane   89 - 105 ( 89 - 105)

----- Final Results -----
      bacterial membrane --- Certainty=0.3590(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
45  bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

>GP:BAB06435 GB:AP001516 two-component sensor histidine kinase
[Bacillus halodurans]
50  Identities = 118/427 (27%), Positives = 199/427 (45%), Gaps = 25/427 (5%)

Query: 10  LERRQRIISAIATA-LAAQINISILADGFIMTSLFLFVFLYFNDDINPILLCLGITF 68
L + II+S + A +A +IN + + F ++L I +FL F + I+
Sbjct: 7   LSKDYMIILSMLLFAPAPIAGEINFYPVNETFRVSLGPPIFFLFLRLNTAAIVPGFFTAI 66

55  Query: 69  ASPIFRGIILSIAGEAEIHQIIEFVLTDMAFYICYGITFYTTYWHRSYRNKGTFFFSIII 128
A ++FR + ++ + E FY Y + F R + F II
Sbjct: 67  AVVVFRVFLDTLHADFYWVDSFEIHYPTFFFYFTYSLLESLAKVQRFHEQPLIIFLFGII 126

```

-1176-

Query: 129 CDYFANLVEISFLIKFNNTITIFA-TLFAIALLRFAFISCAVAYTYSYLSLLLOKD---D 184
 + A+ E F+ ++ + + + ++F I L+ S V +S+ L + + +
 Sbjct: 127 IEILADTAE--FIAQYFAFGVMVTKDSIFQILLIAFSHSFIVLGVFSMMKLYETRSRELE 184

Query: 185 HERRYYYFMWSTSAVKSEVYFMQKNIIIEIENIMKNAYLLDKELSKY--HLPKEYQHLS- 240
 +R + + S + E ++K + E+I + L +E+ + H+ + HL
 Sbjct: 185 IRKRNEHMLLLISNLYEESVHLKKTLONSEDITSKVFLYREMKRLQSEHMDQVNPHEK 244

Query: 241 -----LDISRDVHEVKKDYQNIKGLGTYFSVKNESTMALKDIFQIVLSYTRS---IIQF 292
 L+IS +VHE+KKD Q I GL S NES + +I QI+ R+ Q
 Sbjct: 245 ISKRLLLEISGEVHEIKDNQRIFAGLSKLIS--NESYVDYIRIGQIIKMIVRTNEKYAQL 302

Query: 293 RHQDIILNENKCNLIISNYYYLLTIISNIVLNAVEAIDKQKGTISVHTEEBEDFIKIE 352
 ++I + + + Y L+II+N+V NAVEAID KG +++ + L ++
 Sbjct: 303 LGKEIDFHYSIQGEHPYHYIYTHLSIINNLVANAVEAIDG--KGMLTIRVKALGQTVEFR 360

Query: 353 ISDNGPGIPDKMKHMIKPGFSTKFDANGDIYRGIGLSHVRIIMEBQYQGTITVCPNQ-P 411
 I D+GEGIPDK + +IFKPGF++KFD G GIGL++V M ++ GT+ Q
 Sbjct: 361 IEDDGPICPDKHALIFKPGFTSKFDHTGKPGTGIGLTYVHD-MVDKLGCTVVYVERGQGG 419

Query: 412 NGTTFITL 418
 G+ FT+
 Sbjct: 420 KGSVFTI 426

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1052

A DNA sequence (GBSx1124) was identified in *S.agalactiae* <SEQ ID 3245> which encodes the amino acid sequence <SEQ ID 3246>. This protein is predicted to be ornithine carbamoyltransferase Otc6850 (argF). Analysis of this protein sequence reveals the following:

Possible site: 61
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.64 Transmembrane 171 - 187 (171 - 187)

----- Final Results -----
 bacterial membrane --- Certainty=0.1256(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB75986 GB:AJ272085 ornithine carbamoyltransferase
 [Staphylococcus aureus]
 Identities = 264/332 (79%), Positives = 292/332 (87%)

Query: 1 MKNLRNRSFLTLLDFSTAEVEFLLKLSSEDLKRAKYAGIEQQKLVGKNIALIFEKDSRTR 60
 MKNLRNRSFLTLLDFS EVEFLL LSEDLKRAKY G E+ L KNIAL+FEKDSRTR
 Sbjct: 1 MKNLRNRSFLTLLDFSROEVEFLLTSEDLKRAKYIGTEKPMKKNKIALLFKDSRTR 60

Query: 61 CAFEVAHDQGAHVITYLGPTGSQMGGKETS KDTARVLGGMYDGIERYGFSQETVETLAEF 120
 CAFEVAHDQGA+VTYLGPTGSQMGGKET+KDTARVLGGMYDGIERYGFSQ TVETLAE+
 Sbjct: 61 CAFEVAHDQGANVTYLGPTGSQMGGKETTKDTARVLGGMYDGIERYGFSQRTVETLAEY 120

Query: 121 SGVPVWNGLTADHPTQVLADFLTAKCECLKPYKDIRFTYVGDGRNNVANALMIGASIVG 180
 SGVPVWNGLTADHPTQVLADFLTAK L K Y DI FTYVGDGRNNVANALM GA+I+G
 Sbjct: 121 SGVPVWNGLTDEHPTQVLADFLTAKLEVKKDYADINFTYVGDGRNNVANALMQGAING 180

Query: 181 MTYHLVCPKELEPDPPELLSKQEIATTTGASIEITADIAEGVRDSDVLYTDVWVSMGEPD 240
 M +HLVCPKEL P ELL++C+ IA G +I IT DI +GV+ SDV+YTDVWVSMGEPD
 Sbjct: 181 MNFHLVCPKELNPTDELLNRCKNIAAENGGNILITDDIDQGVKGSVDVIYTDVWVSMGEPD 240

-1177-

Query: 241 EVWKERIALLEPYRITQEMLNMTENPNVIFEHCLPSFHNIDTKVGYDIYEKYGLKEMEVS 300
 EVWKER+ LL+PY++ +EM++ T NPNVIFEHCLPSFHN DTK+G I+EKYG++EMEV+
 Sbjct: 241 EVWKERLELLKPYQVNKEMMDKTGNPNVIFEHCLPSFHNADTKIGQQIFEKYGIREMEVT 300

5 Query: 301 DEVFEGPHSVVFQEAENRMHTIKAVMVATLGD 332
 DEVFE SVVFQEAENRMHTIKAVMVATLG+
 Sbjct: 301 DEVFESKASVVFQEAENRMHTIKAVMVATLGE 332

There is also homology to SEQ ID 3118.

10 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1053

A DNA sequence (GBSx1126) was identified in *S. agalactiae* <SEQ ID 3247> which encodes the amino acid sequence <SEQ ID 3248>. This protein is predicted to be carbamate kinase (b2874). Analysis of this
 15 protein sequence reveals the following:

Possible site: 53
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.48 Transmembrane 214 - 230 (214 - 230)
 20 ----- Final Results -----
 bacterial membrane --- Certainty=0.1192(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

25 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAA66367 GB:X97768 carbamate kinase [Clostridium perfringens]
 Identities = 162/313 (51%), Positives = 207/313 (65%), Gaps = 7/313 (2%)

30 Query: 3 KIVVALGGNAL-----GNSPEEQLRLVKHTAKSLVALIKKGHEIVVSHGNGPQVGAINLG 57
 KIV+ALG NAL S E QL + TA S+ LI+ GHE+ + HGNGPQVG I
 Sbjct: 2 KIVLALGENALQKDSKDKSAEGQLETCRQTALISVADLTIEDGHEVSIVHGNGPQVGQILAS 61
 Query: 58 MNFAAESGQGTN-FPFPECGAMSQGYIGYHLQSSLNELRQEGINKEVATITITQIEVDES 116
 + A + G FPF GA S+GYIGYHLQ ++ EL + GI K V TI TQ+ VD++
 35 Sbjct: 62 IELAHQVDNGNPLFPFDVVGAFSEGYIGYHLQNTIRELLKRGIEKSVDTITTQVIVDKN 121
 Query: 117 DQAFSAPTKPIGTFYDKETSEKIAIEKGYTFVEDAGRGYRRVVASBPBPKKIIINSIKTL 176
 D F+ PTKPIG+FY KE +EK+ +KGYT EDAGRGYRRVVASP+P I+E +IKT+
 Sbjct: 122 DPGFTNPTKPIGSFYTKEEAEKLEKDKGYTMKEDAGRGYRRVVASPKPVDIVEKEAIKTM 181
 40 Query: 177 IENDTLVIAGGGGIPVINKGG-YEGIAAVIDKDKSSALLAGELAADQLIILTAVDYVYT 235
 +++ +VIA GGGGIPV+ G EG+ AVIDKD ++ LA L AD L+ILTAVD V
 Sbjct: 182 VDSGFIVIAACGGGIPVVEDGDRLEGVPAVIDKDFAAEKLAELDADALLILTAVDRVCV 241
 45 Query: 236 QFGKENQKALTEVNENQMIDYVNQGEFAKGSMLPKVIACMSFLDHNPKGTALITSLNGLE 295
 F K +QKAL E+N ++ Y+ +G+FA GSMLPKV AC F+ K A+I SL +
 Sbjct: 242 NFNKPDQKALKEINLEEVDKYIEEGQFAPGSMLPKVEACKKFVLSGDKKVAIIASLTNAK 301
 Query: 296 DALDGKLGTRITK 308
 50 AL G+ GT+I K
 Sbjct: 302 AALRGESGTKIVK 314

There is also homology to SEQ ID 3110.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for
 55 vaccines or diagnostics.

-1178-

Example 1054

A DNA sequence (GBSx1127) was identified in *S.agalactiae* <SEQ ID 3249> which encodes the amino acid sequence <SEQ ID 3250>. Analysis of this protein sequence reveals the following:

```

Possible site: 34
5  >>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3558(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
10     bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1055

A DNA sequence (GBSx1128) was identified in *S.agalactiae* <SEQ ID 3251> which encodes the amino acid sequence <SEQ ID 3252>. This protein is predicted to be a transmembrane protein (b2298). Analysis of this protein sequence reveals the following:

```

20  Possible site: 35
    >>> Seems to have a cleavable N-term signal seq.
      INTEGRAL    Likelihood = -13.11    Transmembrane  413 - 429 ( 405 - 440)
      INTEGRAL    Likelihood =  -9.61    Transmembrane  498 - 514 ( 489 - 516)
      INTEGRAL    Likelihood =  -9.45    Transmembrane  165 - 181 ( 161 - 185)
25  INTEGRAL    Likelihood =  -8.07    Transmembrane  127 - 143 ( 122 - 146)
      INTEGRAL    Likelihood =  -7.22    Transmembrane  308 - 324 ( 306 - 326)
      INTEGRAL    Likelihood =  -5.57    Transmembrane  334 - 350 ( 330 - 357)
      INTEGRAL    Likelihood =  -4.51    Transmembrane  194 - 210 ( 193 - 217)
      INTEGRAL    Likelihood =  -3.82    Transmembrane  372 - 388 ( 371 - 390)
30  INTEGRAL    Likelihood =  -1.22    Transmembrane  250 - 266 ( 250 - 268)
      INTEGRAL    Likelihood =  -0.80    Transmembrane  468 - 484 ( 468 - 484)
      INTEGRAL    Likelihood =  -0.32    Transmembrane  436 - 452 ( 436 - 452)

----- Final Results -----
35     bacterial membrane --- Certainty=0.6243(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database.

```

40  >GP:AAC22251 GB:U32741 conserved hypothetical transmembrane protein
      [Haemophilus influenzae Rd]
      Identities = 303/506 (59%), Positives = 389/506 (75%), Gaps = 6/506 (1%)

Query: 10  NKRSKGRMPGAFTILFILTFISVLATWWIPAGSYSKLQFDTASSKLVTDPNGKTVHVP 69
      +K+ K F P AFTILF + I +V TW IP+GSYSKL +++ + VV P
45  Sbjct: 4  SKKKKTFNFPFAFTILFAILILAVGLTWIPSGSYSKLTYNSTDNVFVVKAYGVDDKTYP 63

Query: 70  ATQTQLDKMNVKIKIKEFTSGAISKPVSVPNTYKRLKQNPAGIGSVTTSMVNGTIEAVDI 129
      AT LD +N+KIK+ FT G I KP+++P TY+R++Q+ GI +T SMV GTIEAVD+
50  Sbjct: 64 ATTDITLDNLNLIKILSNFTEGVIKPIAIPGTYQRVEQHHKGIEDITKSMVEGTIEAVDV 123

Query: 130 MVFIMVLGGMIGVVRKSGAFESGLLALTKKTKGREFLILFLVSLLMVLGGTLCGIEEEAV 189
      MVFI VLGGMIGV+ ++G+F +GL+AL KKTGK EF ++F VS+LMVLGGT CGIEEEAV
Sbjct: 124 MVFIFVLGGMIGVINRTGSFNFAGLMALVKKTKGNEFFIVFCVSVLMVLGGTTCGIEEEAV 183
55  Query: 190 AFYPILVPIFLAMGYDSIIICVGAIFLASSVGTSTFSTINPFSSVIASNAAGISFTEGLSWR 249

```

-1179-

AFYPILVP+FLA+GYD+I+CVGAIFLA+S+GT+ESTINPFS VIASNAAGI FTEG+ +R
 Sbjct: 184 AFYPILVPVFLALGYDAIVCVGAIFLAASMGTAFTSTINPFSVVIASNAAGIQFTEGIGFR 243

Query: 250 TAGCIAGAIFFVVYLHWYAKKIKANPEFSYSYEDRVEFNAKMGMTIN-HTPSLFTIRQKI 308
 G + GA V+ YL+WY KKIK+P FSY+Y+DR EF ++ + +T F+ R+K+
 Sbjct: 244 ALGLVLGATCVIAYLYWYCKKIKADPSFSYTYDDREEFQRQYMKNFDPNTTIPFSARRKL 303

Query: 309 ILSLFVISFPLMVWGVMSQGWFFPTMASSFLAITIIIMFLTATGANGIGERDVVDEFVNG 368
 IL+LF ISFP+M+WGVM GWWFP MA+SFLAITIIIMF+ +G+ E+D+++ F G
 Sbjct: 304 ILTLFCISFPIMIWGVMVGWFFQMAASFLAITIIIMFI-----SGLSEKDIMESFTEG 358

Query: 369 ASSLVGVSLLIIGLARGINIILSQGYISDTMLYTASKLASHVSGSVFIIVMMFIYFVLGVFV 428
 AS LVGVSLLIIGLARG+N++L QG ISDT+L S + S + GSVFI+ + ++ LG +
 Sbjct: 359 ASELVGVSLLIIGLARGVNLVLEQGMISDTILDYMSNVVSGMPGVSFILGQLVVFIFLGLI 418

Query: 429 VPSSSGLAVLSMPILAPLADTVGIPRSVVVMAYQFGQYAMLFAPLPTGLVMATLQMLDMKY 488
 VPSSSGLAVLSMPI+APLAD+VGIPR +VV AY +GQYAMLFAPLPTGLV+ TLQML + +
 Sbjct: 419 VPSSSGLAVLSMPIAPLADSVGIPRDIVVSAYNWQYAMLFAPLPTGLVLVTLQMLQIPF 478

Query: 489 SHWLKFWVPVVLFLIFGGGLLVQLV 514
 W+KfV P++ LL+ G LLV+QV
 Sbjct: 479 DRWVKFVMPMIGCLLLIGSILLVVQV 504

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3253> which encodes the amino acid
 sequence <SEQ ID 3254>. Analysis of this protein sequence reveals the following:

Possible site: 36
 >>> Seems to have a cleavable N-term signal seq.

INTEGRAL	Likelihood = -13.21	Transmembrane	479 - 495 (472 - 496)
INTEGRAL	Likelihood = -10.24	Transmembrane	261 - 277 (258 - 280)
INTEGRAL	Likelihood = -9.24	Transmembrane	153 - 169 (142 - 180)
INTEGRAL	Likelihood = -7.17	Transmembrane	393 - 409 (391 - 411)
INTEGRAL	Likelihood = -6.00	Transmembrane	81 - 97 (78 - 99)
INTEGRAL	Likelihood = -5.95	Transmembrane	318 - 334 (314 - 338)
INTEGRAL	Likelihood = -3.77	Transmembrane	352 - 368 (352 - 369)
INTEGRAL	Likelihood = -2.66	Transmembrane	120 - 136 (119 - 138)
INTEGRAL	Likelihood = -0.32	Transmembrane	204 - 220 (204 - 220)

----- Final Results -----
 bacterial membrane --- Certainty=0.6286(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAB94000 GB:AF008219 unknown [Borrelia afzelii]
 Identities = 174/496 (35%), Positives = 306/496 (61%), Gaps = 37/496 (7%)

Query: 10 RIPSSYTVLFIIIAIMAVLTWFIPAGAYETAK---GGG-----VISGTYKTVASNPQGGFF 61
 ++PSS+T++F +I + +LT+ IPAG ++ G G +++GTY+T+ P+GF
 Sbjct: 3 KMPSSFTIIFSLIVFVTILTYPVIPAGKFDKEFRQIGDGPKREIIVAGTYQTIDRGPRGFL 62

Query: 62 DILMAPVRGMLGVEGTDGAIQVSFFILMVGGFLGVNKTGALDTGIASVVRKNKGREKML 121
 +M + M +G + A +V F+L+VGG G++ KTGA+D GI S+++K ++K+L
 Sbjct: 63 HPIMTILTAMS--KGMEHAAEVIIIFVLIVGGAYGIIMKTGAIDAGIYSLIKKLGHKDKLL 120

Query: 122 IAILIPLFALGGTTYGMGEETMAFYPLLPVMIAVGFDSIVAVAILIGSQIGCLASTIN 181
 I +L+ +F++GGT GM EET+ FY ++IP+++A+G+D++V VAIL +G+ +G +AST+N
 Sbjct: 121 IPLLMFIFSIGGTVTGMSEETLPFYFVMIPLIVALGYDNVVGVAIIALGAGVGTMASTVN 180

Query: 182 PFATGVAADAAGVSIADGMTWRVVIQWVILVGMSIWFVYNYASKIEDPSKSLVADKEEH 241
 PFATG+A+ A +S+ DG +R++ + I + ++I +V YAS+I++DPSKSLV K+ EH
 Sbjct: 181 PFATGIASAIASISLQDGFSPRIVLYFISILVAILIYCVYASRIKKDPSKSLVYSKKNEH 240

Query: 242 KELF-QLQNSGEDLNKRQRNVLTIFTLTFFVIMILSLIPWEDFGIKFTNTINTWLTMPIL 300
 + F + + S ED NV TF ++ L+ FG I + ++ L
 Sbjct: 241 YQYFVKNEISKED-----NVQNTLEFTFARKLVLLL-----FGFM-----ILFLVFSIVQL 286

-1180-

Query: 301 GGVIGKTMGAFGTWYFPEITMLFIMMGVLVAIVYRMSEEDFFSSFLTIGAGEFLGVAMICA 360
 G W+ E+TML++ + ++ A + R+ E + + +F+ G+ + A+I
 Sbjct: 287 G-----WWWQEMTMLYLGVAILSAFTICRLGESEMWDAFVKGSSESLITAALIIG 334

Query: 361 IARGIQVIMNGMITATILHLGETSLSGLSQVFVILAYIFYLPMSEFLIPSTSLAGATM 420
 +ARG+ ++ + G+ITAT+L+ L L F+IL I + + F++PS+SG A TM
 Sbjct: 335 LARGVMIVCDDGLITATMLNAATNFLYNLPRPFFIILNEIIQIFIGFIVPSSSGHASLTM 394

Query: 421 GIMAPLQFQSNVPAHLVITAFQASGILNMISPTSAIVMGALALGRVDLGTWWKFIGKFI 480
 IMAPL F ++ V+ A Q++SG++N+I+PTS ++M L + ++ GTW+KF+
 Sbjct: 395 PIMAPLADFLSIGRSSVVIAMQTSGLINLITPTSGVIMAVLGISKLSYGTWFKFVLPFLF 454

Query: 481 VMVMLVSVLLLVVATF 496
 ++ +S+L+++ +
 Sbjct: 455 IIEFFISILVIANVY 470

An alignment of the GAS and GBS proteins is shown below.

Identities = 158/542 (29%), Positives = 274/542 (50%), Gaps = 92/542 (16%)

Query: 11 KRSGKFRMPGAFTILFILTIFSVLATWWIPAGSYSKLQFDTASSKLVTDPNGKTVHVP 70
 ++ +GFR+P ++T+LFI+ + TW+IPAG+Y +TA
 Sbjct: 4 EKKRGFRIPSSYTVLFIIIAIMAVLTWFIPAGAY-----ETAKG----- 42

Query: 71 TQTQLDKMNVKIKIKEFTSGAISKVSPNTYKRLKQNPAGIGSVTSMVNG-----TI 124
 G IS TYK + NP G + + V G T
 Sbjct: 43 -----GGVIS-----GTYKTVASNPQGFDDILMAPVRGMLGVEGTD 78

Query: 125 EAVDIMVFIMVLGGMIGVVRKSGAFESGLLALTKKTKGREFLILFLVSLMLVGGTLCGI 184
 A+ + FI+++GG +GVV K+GA ++G+ ++ +K KGRE +LI ++ L LGGT G+
 Sbjct: 79 GAIQVSFFILMVGGFLGVVNKTGALDGTGIASVVRKNKGREKMLIAILIPFALGGTTYGM 138

Query: 185 EEEAVAFYPILVPIFLAMGYDSIICVGAIFLASSVGTFSFSTINPFSSVIASNAAGISFTE 244
 EE +AFYP+L+P+ +A+G+DSI+ V I + S +G STINPF++ +A++AAG+S +
 Sbjct: 139 GEETMAFYPLLIPVMIAGVGFDSIVAVAILIGSQIGCLASTINPFATGVAADAAGVSIAD 198

Query: 245 GLSWRTAGCIAGAFVSVYLHWYAKKIKANPEFSYSYEDRVEFNAKWCMTNHTPSLF 304
 G+ WR + + +++ YA KI+ +P S D+ E + + N L
 Sbjct: 199 GMIWRVIQWVILVGMISWIFVNYASKIEEDPSKSL-VADKEEHHKELFQLQNSGEDL-NK 256

Query: 305 RQKIILSLFVISFPLMV-----W-----GVMSQ-----GWWF 331
 RQ+ +L++F ++F +M+ W GV+ + W+F
 Sbjct: 257 RQRNVLTIFTLTFLVIMILSLIPWEDFGIKFFTNINTWLTTPILGGVIGKTMGAFGTWYF 316

Query: 332 PTMASSFLAITIIMFLTATGANGIGERDVVDEFVNGASSLVGVSLIIGLARGINIILSQ 391
 P + F+ + +++ + + E D F+ GA +GV++I +ARGI +I++
 Sbjct: 317 PEITMLFIMMGVLVAIVYR-----MSEEDFFSSFLTIGAGEFLGVAMICAIARGIQVIMNG 371

Query: 392 GYISDTIMLYTASKLASHVSGSVFIIVMMFIYFVLGVFPSSSGLAVLSMPILAPLADTVG 451
 G I+ T+L+ S +S VF+I+ Y + F++PS+SGLA +M I+APL
 Sbjct: 372 GMITATILHLGETSLSGLSQVFVILAYIFYLPMSEFLIPSTSLAGATMGIMAPLQFSN 431

Query: 452 IPRSVVVMAYQFGQYAMLEFLAPT-GLVMATLQMLDMKYSHWLKFVWPVVLFLIFGGGLLV 512
 +P +V+ A+Q + ++PT +VM L + + W KF+ ++ +++ LLV+
 Sbjct: 432 VPAHLVITAFQASGILNMISPTSAIVMGALALGRVDLGTWWKFIGKFIIVMMLVSVLLLV 493

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1056

A DNA sequence (GBSx1129) was identified in *S.agalactiae* <SEQ ID 3255> which encodes the amino acid sequence <SEQ ID 3256>. Analysis of this protein sequence reveals the following:

Possible site: 46

-1181-

>>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood =-10.83 Transmembrane 25 - 41 (18 - 47)
 INTEGRAL Likelihood =-10.46 Transmembrane 153 - 169 (148 - 176)

5 ----- Final Results -----
 bacterial membrane --- Certainty=0.5331(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

10 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB13183 GB:Z99110 similar to two-component sensor histidine
 kinase [YkoG] [Bacillus subtilis]
 Identities = 119/446 (26%), Positives = 212/446 (46%), Gaps = 18/446 (4%)

15 Query: 17 TQITLWYSSFIFILVIGVLIGSFFISKSAENKSKKNLEAKAVQMSQALAKGHRYEAFED 76
 T+I L+ S + IL+I V + I S +K L + +++AL
 Sbjct: 5 TKIHLTYSISLLILLILVHTAVYLIFSSALTSKDAARLADETDNIAEALRAAETEGVALQ 64

20 Query: 77 GIFYSVYDQNGKV-IYSGFPGKFKRDLDDHQHKKHKKLSLFSMEN-----RTFCYVDI 127
 + + NG V + +G K + LS S E + F +
 Sbjct: 65 DMLQAYLPANGMVRVVGNDQKAVMTITKEKAYKDFPLSFHSGETADVVRKPDGKLF AEAAV 124

25 Query: 128 PISGNQWLRAIRTVDRLDKQLTEFLSLGIVLPLMLIIITVG---GYLILKRTFRPIQ 183
 P+ + + + + V+RL+ E LF L I+L + + G L+ +R PI+
 Sbjct: 125 PVIWTDGQVVSLLQVLERLENT-EESLFLKIIILIAASAACIASFFAGSLLARRIINPIR 183

30 Query: 184 EITETAQFITQNEYTKRIITKNENELTELAAVINTMLASIESFVREKQFNNDVSHEL 243
 + T + I +++++ + + +EL ++ N M ++ + + + +QF D SHEL
 Sbjct: 184 RLMITMKDIQRDKFETISLEGQSNDELYQMGLTFNEMAMMLKEHYDKQQQFVQDASHEL 243

35 Query: 244 RTPVTVILSESEYGNKYAENLSEA-KESFEVIHRQSLSMKKLVEQLLELTKAENPLSIQL 302
 +TP+T+I S S K + E +ES E IH +++ MKKL QLL L K+ L + L
 Sbjct: 244 KTPLTIIESYSSLMKRWGAKKPEVLEESIEAIHSEAVHMKKLTNQLLALAKSHQGLEVDL 303

40 Query: 303 EPLNFSIMMKQLVSDSSRLDNTPIHLDSQIEDDLWIIGQQTLLKRLFDNLFNSNAIKFTN 362
 + + + I + V + + + I L++ ++ L + + +K+L L NAIK++
 Sbjct: 304 KTIDL- IKAARAVMQTLQSVYQRDILLETD-KESLLVKADEERIKQLLTILLDNAIKYSE 361

45 Query: 363 NHISISLRQSDNQIVFSIKDNLGLISVDDQSKIWNRFYQVDSARTKDSQSGIGLGLSLVK 422
 I +S + + S++D G+GI + ++ RFY+ D AR + + G GLGLS+ K
 Sbjct: 362 KPIEMSAGTRNGRPFLSVRDEGIGIPEEHIPHLFERFYRADEARNRKT-GGTGLGLSLIAK 420

Query: 423 QIATIHRAKIWVDSKPDGSGQFTLTF 448
 QIA H ++ V SKP G+ T+ F
 Sbjct: 421 QIADEHGIELSVKSKPGQGTAVTMQF 446

There is also homology to SEQ ID 1178.

SEQ ID 3256 (GBS77) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 21 (lane 2; MW 78.5kDa) and in Figure 28 (lane 2; MW 78.5kDa).

50 GBS77-GST was purified as shown in Figure 195, lane 4.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1057

55 A DNA sequence (GBSx1130) was identified in *S.agalactiae* <SEQ ID 3257> which encodes the amino acid sequence <SEQ ID 3258>. This protein is predicted to be CopR protein (tcrA). Analysis of this protein sequence reveals the following:

Possible site: 33

-1182-

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.3963(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

10 >GP:CAC07978 GB:AJ278983 CopR protein [Ralstonia metallidurans]
 Identities = 102/221 (46%), Positives = 145/221 (65%)

Query: 1 MKILVVEDEFDLNRISIVKLLKKQHYSVDSASNGEEALQFVSVAEYDVIIIDVMMPKMDGF 60
 MK+LVVEDE + + L + + VD +NG + F YD+IILDVM+P +DG+
 15 Sbjct: 1 MKLLVVEDEVKTGEYLRQGLTEAGFVVDLVANGLDGQHFAVNETYDLIIIDVMLPDVDGW 60

Query: 61 TFLKLLRNKGSQVSILMLTARDAVEDRIAGLDFGADDYLVKPFEGELMARIRAMLRRAN 120
 L +R G+ V +L LTARD+V DR+ GL+ GADDYLVKPF F EL+AR+R +LRR
 Sbjct: 61 HILHAIRASGNAPVFLTLTARDSVADRVRGLELGADDYLVKPFAPSELLARVRTLLRGA 120

20 Query: 121 RQVSSDDIQIDITINLSTKQVWRNDNLIDLTAKEYEVLEYLARHRDQVLSRQIREHWV 180
 Q++ D IQ+ D+ ++LS ++ R I LT+KE+ +LE AR R +VL R I VW
 Sbjct: 121 VOLAMDRIQVADLILDLRRLRRASRGRRITLTSKEFALLELFARRRGEVLPRLIASQVW 180

25 Query: 181 DYDYGESNIIDVLIKNLRRKLDNNRDGSLIKTKRGLGYVI 221
 D ++ +SN+IDV I+ LR K+D+ + LI+T RG+GYV+
 Sbjct: 181 DMNFDSNVIDVAIRRLRAKIDGFEVKLIQTVRGMGYVL 221

There is also homology to SEQ ID 3260.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for
 30 vaccines or diagnostics.

Example 1058

A DNA sequence (GBSx1131) was identified in *S.galactiae* <SEQ ID 3261> which encodes the amino acid sequence <SEQ ID 3262>. Analysis of this protein sequence reveals the following:

35 Possible site: 40
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -3.45 Transmembrane 18 - 34 (16 - 36)

----- Final Results -----

40 bacterial membrane --- Certainty=0.2381(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10281> which encodes amino acid sequence <SEQ ID 10282> was also identified.

45 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 3262 (GBS78) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 16 (lane 4; MW 23.8kDa).

50 The GBS78-GST fusion product was purified (Figure 194, lane 4) and used to immunise mice. The resulting antiserum was used for FACS (Figure 317), which confirmed that the protein is immunoaccessible on GBS bacteria.

-1183-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1059

A DNA sequence (GBSx1132) was identified in *S.agalactiae* <SEQ ID 3263> which encodes the amino acid sequence <SEQ ID 3264>. Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have an uncleavable N-term signal seq
    INTEGRAL    Likelihood = -11.04    Transmembrane    15 - 31 ( 6 - 35)
    INTEGRAL    Likelihood = -1.28    Transmembrane    51 - 67 ( 51 - 67)

----- Final Results -----
    bacterial membrane --- Certainty=0.5416(Affirmative) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 3264 (GBS79) was expressed in *E.coli* as a GST-fusion product. GBS79d was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 154 (lane 17 & 18; MW 51kDa), in Figure 155 (lane 17; MW 51kDa) and in Figure 187 (lane 13; MW 51kDa). GBS79d was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 155 (lane 2-4; MW 26kDa) and in Figure 183 (lane 5; MW 26kDa). Purified GBS79d-GST is shown in Figure 243, lane 2.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1060

A DNA sequence (GBSx1133) was identified in *S.agalactiae* <SEQ ID 3265> which encodes the amino acid sequence <SEQ ID 3266>. Analysis of this protein sequence reveals the following:

```
Possible site: 50
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
    bacterial cytoplasm --- Certainty=0.5326(Affirmative) < succ>
    bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10279> which encodes amino acid sequence <SEQ ID 10280> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

```
>GP:AAG20974 GB:AE005164 Vng6349c [Halobacterium sp. NRC-1]
Identities = 97/358 (27%), Positives = 163/358 (45%), Gaps = 20/358 (5%)

Query: 35 DPQIIKLTTRANIAIGTYEGFLESIIINPMLLISPLLSQEAVALSSKLEGTHATLKDLLNVE 94
          D + A +G G + P +L + LL +EA+ S+++EG L + E
Sbjct: 70 DDDFYETLADATFWLGLSGVSLDLDFPPVLYTSLLRKEAMESARIEGADVVDYDALYSLE 129

Query: 95 AGNKVDIERDELHEII-----NYRKALFYALENISTINNIDSKGLPLSNRIIKEMHKIL 148
          D RDE E + R+ L Y I+ +D+ G L+ ++ ++H+ L
```

-1184-

Sbjct: 130 T-RTFDEGRDEPSETTAAAEKDTREVLNYETAVKEGIDALDA-GEELNVELLHDLHETL 187

Query: 149 LDNV--RGSSKNPGNFKRSQNYIGSVSSISYTPVPAEKTPEYMSNLEQYIHYD-DLDDL 204
 L V R + G++K + NY+G + P + M L Y L

5 Sbjct: 188 LTGVPPDRVDTDITGDYKINPNYLG-----FLPPAGAVEDLMDGLFTYYRTGGSYHPL 242

Query: 205 VQSALHAQFEMIHPPFEDGNGRIGRLLIPLFLYQELLSYPTFYMSYFERDRSLYISHL 264
 V A+ H QFE IHP+ DGNGR+GRLLI L LY +LL P Y+S Y R+++ Y+ +

10 Sbjct: 243 VDIALFHYQFETIHPYGDGNGRIGRLLITLQLYDADLLERPNLYLSEYLNRNKTTYVERM 302

Query: 265 SNISKDNWWDWFEYYLEGVILSAERSTKKAQDILSLYNIMKEQVIPKLSVSGIQLLDF 324
 + W+ W +++EG+ A ES ++ + L + + K + + QL

15 Sbjct: 303 EGVRFHGEWEAWLSFFIEGIARQAHSVERTRALADLRREYEHEYGGKAYTKN--QLAVT 360

Query: 325 IFSAPIFKAEQVSEHLKISKRTTYTLNKLIDEGYL-STDNAQRNRTYYCPQLLSIVQ 381
 +F P ++ V I + T +N+L++EG L RN+ Y ++ I++

Sbjct: 361 LFEQPYITSKTVQRLFDIEQSTASRAINELVNEGILEEVPRHGRNKEYRAREIFEILE 418

No corresponding DNA sequence was identified in *S.pyogenes*.

- 20 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1061

A DNA sequence (GBSx1134) was identified in *S.agalactiae* <SEQ ID 3267> which encodes the amino acid sequence <SEQ ID 3268>. Analysis of this protein sequence reveals the following:

25 Possible site: 47
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4370(Affirmative) < succ>
 30 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

RGD motif : 46-48

- 35 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 3268 (GBS299) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 58 (lane 2; MW 62.2kDa) and in Figure 60 (lane 4; MW 62.2kDa).

GBS299-GST was purified as shown in Figure 207 (lane 4) and Figure 225 (lanes 2-3).

- 40 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1062

A DNA sequence (GBSx1135) was identified in *S.agalactiae* <SEQ ID 3269> which encodes the amino acid sequence <SEQ ID 3270>. Analysis of this protein sequence reveals the following:

45 Possible site: 37
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4176(Affirmative) < succ>
 50 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1063

A DNA sequence (GBSx1136) was identified in *S.agalactiae* <SEQ ID 3271> which encodes the amino acid sequence <SEQ ID 3272>. Analysis of this protein sequence reveals the following:

```

Possible site: 19
10  >>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.1789(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
15      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1064

A DNA sequence (GBSx1137) was identified in *S.agalactiae* <SEQ ID 3273> which encodes the amino acid sequence <SEQ ID 3274>. Analysis of this protein sequence reveals the following:

```

Possible site: 49
25  >>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3748(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
30      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1065

A DNA sequence (GBSx1138) was identified in *S.agalactiae* <SEQ ID 3275> which encodes the amino acid sequence <SEQ ID 3276>. Analysis of this protein sequence reveals the following:

```

Possible site: 51
40  >>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.1638(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
45      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

-1186-

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB12294 GB:Z99106 similar to transposon protein [Bacillus subtilis]
Identities = 84/291 (28%), Positives = 138/291 (46%), Gaps = 6/291 (2%)

5 Query: 6 MLDYLAVTIKGLAPDDVIEKILILPKDKFVLNEWGINKYQRHYSFSEIKVYFNKDWQSKM 65
M+DY+ V+ K D +IE++L L KD + G Y Y IKV+++ ++
Sbjct: 31 MVDYIRVSFKTHDVRDRIIEEVLHLSKDFMTEKQSGFYGVGTIELDYIKVFYSAPDDNR- 89

10 Query: 66 GVFIELRGQGCRCQYEEYMENNWNWVTLMKRISECHSNVTRLDIANDIFDDSLSVPLIYS 125
GV IE+ GQGCRCQ+E ++E W + + + TR D+A D S+P +
Sbjct: 90 GVLIEMSGQGCRCQFESFLECRKKTWYDFQDCMQGGSFTRFDLAIDDKKTYFSIPELLK 149

15 Query: 126 YCKKQLCISTAKTFDYHEKSLLENGKVGEMVTIGVRGTQQW-CVYNKLEQLDQELPN 184
+K CIS + D++ L +G G + G + ++ + C Y K EQ +P
Sbjct: 150 KAKQGECSIRFRKSDFNCSFDLSDGITGTTIYFGSKKSEAYLCFYEKNYEQAEKYNIP 209

20 Query: 185 TPL-SWTRAEELRCWQEKANLLAKQIEGRPLKEIYFEVINGHYRFVSPRDKDSNRWRRT 243
L W R ELR E+A + + + + L I ++IN + RFV D++ R KT
Sbjct: 210 EELGDMNRYELRLKNERAQAIDALLKTKDLTLIAMQIINNYVRFVD-ADENITREHWKT 268

25 Query: 244 VKWWDYLETQEKTVLSVKRTKPTLKRSEKWTEKQVSRITLGLYVAKAESH 294
+W+D++ + L VK K ++S W + T+ V +A+ H
Sbjct: 269 SLFWSDFIGDVGRPLPVKPKDFYQKSRNWLNRNSCAPTM--KMVLEADEH 317

25 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 1066

30 A DNA sequence (GBSx1139) was identified in *Sagalactiae* <SEQ ID 3277> which encodes the amino acid sequence <SEQ ID 3278>. This protein is predicted to be integrase. Analysis of this protein sequence reveals the following:

Possible site: 58
>>> Seems to have no N-terminal signal sequence

35 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.1914(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

40 The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB70622 GB:AJ243106 integrase [Streptococcus thermophilus]
Identities = 135/474 (28%), Positives = 233/474 (48%), Gaps = 68/474 (14%)

45 Query: 20 KAGNVLVKFAMRFTHTPKKSHKKYLSTGASKGWFTTKATPSKKLPSGKERLLVSDIKNT 79
K G + VKF F + +T K ++ LS W+T +KK +GK +L S
Sbjct: 19 KTGIEVVKFRITYFNQLINK-RREILSD-----WYTIV---NKKDTTGKIKL--SPQIKA 67

50 Query: 80 QLITQVTQELNKLVDYIAELMGIKPKAKKLLTLEETAKPFDKDGNYGKAFKAWH--- 136
+ ++ ++ NK+ ++ ++ K +TL+E+ + WH
Sbjct: 68 LIHKELQEKANKVYBELRTIL-----LEKSDITLDEV-----WNEWHNER 108

55 Query: 137 -ERVKPANNTLKRVTIYNYRIEYENFDTRMSITKFAFMTDEIQNLIN-----ASSMHMAR 190
ER A TL Y +I + SI K + I+NL++ + +A+
Sbjct: 109 VERQLVAPKTLAGEDGRYRNHITKQIP-KNSILK-NIPSSLIKLLDNLYPIGNHKRLAQ 166

Query: 191 NLHIYLMKIFDWSVENGQITLTQDPIASNKVKRRVLTKEEQDK-KREDIAEKYLEASEV 249
+ L I+ +++ + I+ Q+P+ + R+ L S+E D+ K+ DI ++YLE+ E+
Sbjct: 167 GVKSDLTSTYKFAILHDYISPDQNPMPYISIGRKGL--SDELDRLKKSDIEDQYLESWEL 224

-1187-

Query: 250 NHVLRRLIESWTNRPDNLQIADVLRMIFLTGMRPSEVLGLNEDMLDFEKKWIKVHWQRASK 309
 VL ++ + N+ A + LTGMR EVLGL E+ +DF K V RA+
 Sbjct: 225 KEVLSIVRKY-----NEQYARIFEQALTGMRIGEVLGLKEEAIDFNKNIASVIRTRATH 279

5 Query: 310 NKSDMMMEALNLDEKERYRADLTKESVRTIPMSPEVEKILRHYIDRNKFQAQFSPTYQD 369
 + + + Y ++K +S R + +S +IL+ I+ N +F+P Y+D
 Sbjct: 280 GCASE-----DSYEGNVKNLQSYRNVQLSKRAIEILKEEIELNHQHIFRNPDKYD 329

10 Query: 370 LGYLFTRTYIRAGNRQGSPLYHNELSQLRGGSSQSAKYNKKAGKPYK---DIDSFLDFG 426
 G++PT I + G+PL+++ L+ FL SS++ K N+ G P + DID+ L F
 Sbjct: 330 NGWIFTSKSIHKPDYNGTPLHYSVLNNFL--NBSSENGKLNRN-GNERRAGIDIDNKLSEFK 386

15 Query: 427 RPIHVIPHFRHSFISIMASEGIDLPTIREFVGHSSEDSKEIERVYLHVIKKQKD 480
 + H+ H+FRH+ IS +A +G+ L I++ VGHS S+ + +YLH+ KK KD
 Sbjct: 387 K--HITTHIFRHTHISFLAEQGVPLEAIQDRVGHSRGSR-VTEIYLHITKTKD 437

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3279> which encodes the amino acid sequence <SEQ ID 3280>. Analysis of this protein sequence reveals the following:

Possible site: 33
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.5203(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 82/357 (22%), Positives = 155/357 (42%), Gaps = 52/357 (14%)

30 Query: 135 WHERVKPANNTLKTTRVTIYNRVIEPNFDTRMSITKFAFMTDEIQNLINA--SSMHMARNL 192
 W K +T + R + D + I K T +Q++I+ S +
 Sbjct: 73 WEHHQKSLKSTSVRSLDFRIEELRNLDPEVMIKIT--TKYLQSIIDKIPGSYDKRKRA 130

35 Query: 193 HIYLMKIFDWSVENGQITLTQDPIASNKVKRRVLTKEEQDKKREDIAEKYLEASEVNVH 252
 LK FD+++ +++ +P+ S ++ + V T K ED+A+K+LE E+
 Sbjct: 131 RQLLQKQTFDYAIALEYVSI--NPVISTQLAKPVKTI-----KDFEDVAQKFLKDELK-- 181

40 Query: 253 LRLIESWTNRPDNLQIADVLRMIFLTGMRPSEVLGLNEDMLDFEKKWIKVHWQRASKNKS 312
 RL++ R + +A + + L G R E L + D + + I++H
 Sbjct: 182 -RLDEMYRRKGSIKMAYLAEFMSLNGCRIGEALAIQPD--NIKNDIIEIH----- 229

45 Query: 313 DDMMEALNLDEKERYRADLTKESVRTIPMSPEVEKILRHYIDRNKFQAQFSPTYQDLGY 372
 ++ + + + KT S R ++ ++I++ + N + +P Y+D+GY
 Sbjct: 230 -GTLDYTSNGYRNAIKTTPKTNSSWRETLLITKREKETIQDILKINALEKNTNPNYKDMGY 288

50 Query: 373 LFTRTYIRAGNRQGSPLYHNELSQLRGGSSQSAKYNKKAGKPYKDIDSFLDFGRPIHVI 432
 +F +R G P+ N L+ +R NK+ KP + +
 Sbjct: 289 IFI-----SRNGVPIQDNALNTSIRAA-----NKRLEKPIQK-----ELT 323

55 Query: 433 PHMFRHSFISIMASEGIDLPTIREFVGHSSEDSKEIERVYLHVIKKQKDTMRGAVEKL 489
 H+FRH+ +S +A + L TI + VGH+ DSK +++Y HV K K+ + + +L
 Sbjct: 324 SHIFRHTLVSRLEANKVPLKTIMDRVGH-DSKTTQQIYTHVTKSMKNEVVDILNRL 379

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1067

A DNA sequence (GBSx1140) was identified in *S.agalactiae* <SEQ ID 3281> which encodes the amino acid sequence <SEQ ID 3282>. Analysis of this protein sequence reveals the following:

Possible site: 42
 >>> Seems to have no N-terminal signal sequence

-1188-

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.3023(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 10277> which encodes amino acid sequence <SEQ ID 10278> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

10 >GP:AAB64982 GB:U43834 Ydr540cp [Saccharomyces cerevisiae]
 Identities = 88/170 (51%), Positives = 117/170 (68%), Gaps = 3/170 (1%)
 Query: 36 MRTYSDKNELKEEVLKSYKKYIAEFNDIPEKLKDLRIDEVDRTPAENLAYQVGWTTLILK 95
 MR Y+ K ELKEE+ K Y+KY ABF I E KD +++ VDRTP+ENL+YQ+GW L+L+
 15 Sbjct: 1 MREYTSKKELKEEIEKKYKYDABFETISESQKDEKVETVDRTPSENLSYQLGWNLLLE 60
 Query: 96 WESDEQSGLEVKTPTETFKWNQLGELYQHFTETYASLTIKELTAQLNDNVDAIGNMIDSM 155
 WE+ E +G V+TP +KWN LG LYQ F + Y +IKE A+L + V+ + I ++
 Sbjct: 61 WEAKEIAGYNVETPAPGYKWNINLGGLYQSFYKKYGIYSIKEQRAKLREAVNEVYKWISTL 120
 20 Query: 156 SDEVLFKPHMRNWADSATKNVAVWEVYKFIHINTVAPFGTFRKIRKWKVKV 205
 SD+ LF+ R W AT A+W VYK+IHINTVAPF FR KIRKWK++
 Sbjct: 121 SDELFLQAGNRKW---ATTKAMWPVYKWIHINTVAPFTNFRGKIRKWKRL 167

25 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1068

30 A DNA sequence (GBSx1141) was identified in *S.agalactiae* <SEQ ID 3283> which encodes the amino acid sequence <SEQ ID 3284>. This protein is predicted to be 50S ribosomal protein subunit L33-related protein. Analysis of this protein sequence reveals the following:

Possible site: 46
 >>> Seems to have no N-terminal signal sequence

35 ----- Final Results -----

 bacterial cytoplasm --- Certainty=0.5420(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

40 The protein has homology with the following sequences in the GENPEPT database.

 >GP:AAB66692 GB:U89998 50S ribosomal protein subunit L33
 [Lactococcus lactis subsp. cremoris]
 Identities = 43/49 (87%), Positives = 46/49 (93%)
 45 Query: 1 MRVNITLEHKESGERLYLT SKNKRNTPDRLQLKKYSPKLRKHVVFTEVK 49
 MRVNITLEHKESGERLYLT KNKRNTPD+L+LKYS KLRKHV+F EVK
 Sbjct: 1 MRVNITLEHKESGERLYLTQKNKRNTPDKLELKKYSKKLRKHVIFKEVK 49

50 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3285> which encodes the amino acid sequence <SEQ ID 3286>. Analysis of this protein sequence reveals the following:

Possible site: 46
 >>> Seems to have no N-terminal signal sequence

55 ----- Final Results -----

 bacterial cytoplasm --- Certainty=0.5394(Affirmative) < succ>

-1189-

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

5 Identities = 48/49 (97%), Positives = 48/49 (97%)
 Query: 1 MRVNITLEHKESGERLYLTSKNKRNTPDRLQLKKYSPKLRKHVVFTFEVK 49
 MRVNITLEHKESGERLYLTSKNKRNTPDRLQLKKYSPKLRKHV FTEVK
 10 Sbjct: 1 MRVNITLEHKESGERLYLTSKNKRNTPDRLQLKKYSPKLRKHVTFTEVK 49

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1069

15 A DNA sequence (GBSx1142) was identified in *S.agalactiae* <SEQ ID 3287> which encodes the amino acid sequence <SEQ ID 3288>. This protein is predicted to be 50S ribosomal protein subunit L32-related protein. Analysis of this protein sequence reveals the following:

Possible site: 56
 >>> Seems to have no N-terminal signal sequence

20 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3577(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

25 The protein has homology with the following sequences in the GENPEPT database.

>GP:AAB66691 GB:U89998 50S ribosomal protein subunit L32
 [Lactococcus lactis subsp. cremoris]
 Identities = 44/53 (83%), Positives = 48/53 (90%)

30 Query: 1 MAKPARHTSKAKRNKRTHYKLTAPSVQFDETTGDYSRSHRVSLKGYKGRKI 53
 MA PARHTS AK+N+RRTHYKLTAP+V FDETTGDY SHRVSLKGYKGRK+
 Sbjct: 1 MAVPARHTSSAKKNRRRTHYKLTAPTVTFDETTGDYRSHRVSLKGYKGRKV 53

35 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3289> which encodes the amino acid sequence <SEQ ID 3290>. Analysis of this protein sequence reveals the following:

Possible site: 35
 >>> Seems to have no N-terminal signal sequence

40 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.5148(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

45 Identities = 38/39 (97%), Positives = 39/39 (99%)
 Query: 22 LTAPSVQFDETTGDYSRSHRVSLKGYKGRKIAKANEAK 60
 +TAPSVQFDETTGDYSRSHRVSLKGYKGRKIAKANEAK
 50 Sbjct: 1 MTAPSVQFDETTGDYSRSHRVSLKGYKGRKIAKANEAK 39

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-1190-

Example 1070

A DNA sequence (GBSx1144) was identified in *S.galactiae* <SEQ ID 3291> which encodes the amino acid sequence <SEQ ID 3292>. This protein is predicted to be histidyl-tRNA synthetase (hisS). Analysis of this protein sequence reveals the following:

```

5   Possible site: 32
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
    bacterial cytoplasm --- Certainty=0.4357(Affirmative) < succ>
10   bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 10275> which encodes amino acid sequence <SEQ ID 10276> was also identified.

15 The protein has homology with the following sequences in the GENPEPT database.

```

    >GP:CAA78919 GB:Z17214 histidine--tRNA ligase [Streptococcus
      equisimilis]
      Identities = 327/404 (80%), Positives = 361/404 (88%)

20   Query: 32  WQYVENVIRNLFKQYHYDEIRTPMFEHYEVISRSVGDITTDIVTKEMYDFHDKGDRHITLR 91
      WQYVE V R FKQYHY *EIRTPMFEHYEVISRSVGDITTDIVTKEMYDF+DKGDRHITLR
      Sbjct: 1   WQYVEGVARETFKQYHYGEIRTPMFEHYEVISRSVGDITTDIVTKEMYDFYDKGDRHITLR 60

      Query: 92  PEGTAPVVRSYVENKLFAPQVQKPTKMYIIGSMFRYERPOAGRLREFHQVGECEFGSNNP 151
      PEGTAPVVRSYVENKLFAPQVQKP K+YYIGSMFRYERPOAGRLREFHQ+GVECEFGS NP
25   Sbjct: 61  PEGTAPVVRSYVENKLFAPQVQKPKLYIIGSMFRYERPOAGRLREFHQIGVECEFGSANNP 120

      Query: 152 ATDVETIAMGHHLFEDLGIKNVKHLHNSLGNPESRQAYRQALIDYLTPIREQLSKDSQRR 211
      ATDVETIAM +HLFE LGIK V LHLNSLGN SR AYRQALIDYL+P+R+ LSKDSQRR
30   Sbjct: 121 ATDVETIAMAYHLFERLGIGVTLHLNSLGNAAASRAAYRQALIDYLSMPMDTLKDSQRR 180

      Query: 212 LNENPLRVLDSKEPEDKLAVENAPSILDYLDSSQAHFDAVCHMLDALNIPYIIDTNMVR 271
      L+ENPLRVLDSKE EDK+AV NAPSILDY DE SQAHFDAV ML+AL IPY+IDTNMVR
35   Sbjct: 181 LDENPLRVLDSKEKEDKIAVANAPSILDYQDEESQAHFDAVRSMLAALAIPLYVIDTNMVR 240

      Query: 272 GLDYNNHTIFEFITEIEDNELTICAGGRYDGLVSYFGGPETPAFGFGLGLERLLLLILDQK 331
      GLDYNNHTIFEFITE++ +ELTICAGGRYDGLV YFGGP TP FGFGLGLERLLLLILDQK
      Sbjct: 241 GLDYNNHTIFEFITEVDQSELTICAGGRYDGLVEYFGGPATPGFGFGLGLERLLLLILDQK 300

40   Query: 332 GISLPIENTIDLYIAVLGSEANLAALDLAQSIHQGFVKVERDYLGKRIKAQFKSADTFNA 391
      G+ LP+E +D+YIAVLG++AN+AAL L Q+IR QGF VERDYLGKRIKAQFKSADTF A
      Sbjct: 301 GVLEPVEEGLDVYIAVLGADANVAALALTQAIRRQGFVERDYLGKRIKAQFKSADTFKA 360

      Query: 392 KVIMTLGSSEVDSKEVGLKNNQTRQEVKVSFENIKTDFSSVLKQ 435
      KV++TLG SE+ + + LK+NQTRQE+ VSF+ I+TDF+S+ +
45   Sbjct: 361 KVVITLGESEIKAGQAVLKHNQTRQEMTVSFDQIQTFASIFAE 404

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3293> which encodes the amino acid sequence <SEQ ID 3294>. Analysis of this protein sequence reveals the following:

```

50   Possible site: 27
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
    bacterial cytoplasm --- Certainty=0.3183(Affirmative) < succ>
55   bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below.

Identities = 339/424 (79%), Positives = 387/424 (90%)

-1191-

Query: 13 MKLQKPKGTQDILPGESAKWQYVENVIRNLFKQYHYDEIRTPMFEHYEVISRSVGDITDI 72
 Sbjct: 1 MKLQKPKGTQDILPGDAAKQYVESVARDTFSQYNYGEIRTPMFEHYEVISRSVGDITDI 60

Query: 73 VTKEYMDFHDKGDRHITLRPEGTA PVVRSYVENKLFAPQVQKPTKMYIIGSMFRYERFQA 132
 Sbjct: 61 VTKEYMDFHDKGDRHITLRPEGTA PVVRSYVENKLFAPQVQKPTKMYIIGSMFRYERFQA 120

Query: 133 GRLREFHQVQVECFGSNNPATDVETIAMGHHLFEDLGIKNVKLHLNSLGNPESRQAYRQA 192
 Sbjct: 121 GRLREFHQVQVECFGSNNPATDVETIAMGHHLFEDLGIKNVKLHLNSLGNPESRQAYRQA 180

Query: 193 LIDYLTPIREQLSKDSQRRLENPLRVLDSKEPEDKLAVENAPSILDYLDSESAHFDAV 252
 Sbjct: 181 LIDYLTPIREQLSKDSQRRLENPLRVLDSKEPEDKLAVENAPSILDYLDSESAHFDAV 240

Query: 253 CHMLDALNIPYIIDTNMVRGLDYYNHTIFEFTIEDNELTICAGGRYDGLVSYFGGPET 312
 Sbjct: 241 CHMLDALNIPYIIDTNMVRGLDYYNHTIFEFTIEDNELTICAGGRYDGLVSYFGGPET 300

Query: 313 PAFGFGGLGLERLLILDKQGISLPIENTIDLYIAVLGSEANLAALDLAQSIHQGFVKVER 372
 Sbjct: 301 PAFGFGGLGLERLLILDKQGISLPIENTIDLYIAVLGSEANLAALDLAQSIHQGFVKVER 360

Query: 373 DYLGRKIKAQFKSADTFNAKIVMTLGSSEVDSKEVGLKNNQTRQEVKVSFENIKTDFSSV 432
 Sbjct: 361 DYLGRKIKAQFKSADTFNAKIVMTLGSSEVDSKEVGLKNNQTRQEVKVSFENIKTDFSSV 420

Query: 433 LKQL 436
 Sbjct: 421 SEQL 424

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1071

A DNA sequence (GBSx1145) was identified in *S. agalactiae* <SEQ ID 3295> which encodes the amino acid sequence <SEQ ID 3296>. This protein is predicted to be aspartyl-tRNA synthetase (aspS). Analysis of this protein sequence reveals the following:

Possible site: 29
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm	---	Certainty=0.5124(Affirmative)	< succ>
bacterial membrane	---	Certainty=0.0000(Not Clear)	< succ>
bacterial outside	---	Certainty=0.0000(Not Clear)	< succ>

A related GBS nucleic acid sequence <SEQ ID 10273> which encodes amino acid sequence <SEQ ID 10274> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB14714 GB:Z99118 aspartyl-tRNA synthetase [Bacillus subtilis]
 Identities = 339/585 (57%), Positives = 432/585 (72%), Gaps = 9/585 (1%)

Query: 20 RSMYAGRVRSEHIGTSTITLKGWVGRRRDLGGLIFIDLRDREGIMQLVINPEEVSASVMAT 79
 Sbjct: 4 RTYYCGDITEKAIGESVTLKGWVQKRRDLGGLIFIDLRDRTGIVQVVFNPV-VSKEALAI 62

Query: 80 AESLRSEFVIEVSGVVTAREQA--NDNLPTGEVELKVOELSILNTSKTTPFEIKDGIE-A 136
 AE +R+E+V+++ G V ARE+ N NL TG +E+ +++LN +KT PF I D E

-1192-

Sbjct: 63 AEGIRNEYVLDIQGKVVAREEGTVNPNLKTGAIEIHADGVNVLNAAKTPPFAISDQAEV 122

Query: 137 NDDTRMRYRYLDLRRPEMLENFKLRAKVTHSIRNYLDNLEFIDVETPMLTKSTPEGARDY 196
 ++D R+++RYLDLRRP M + +LR VT ++R++LD F+D+ETP+IT STPEGARDY

5 Sbjct: 123 SEDVRLKHYRLDLRRPAMFQIMQLRHNVTKAVRSFLDENGFLDIETPILTGSTPEGARDY 182

Query: 197 LVPSRVNQGHFYALPQSPQITKQLLMNAGFDRIYQIVKCFRDEDLRGDRQPEFTQVDLET 256
 LVPSRV++G FYALPQSPQ+ KQLLM +G +RYYQI +CFRDEDLR DRQPEFTQ+D+E

10 Sbjct: 183 LVPSRVHEGEFYALPQSPQLFKQLLMVSGIERYYQIARCFRDEDLRADRQPEFTQIDIEM 242

Query: 257 SFLSDQEIQDIVEGMIKVMKDTKGLEVSLEPFRMAYDDAMNNYGSDDKPDTRFDMLLQDL 316
 SF+S ++I + E M+AKVM++TKG E+ LP PRM YD+AMN YGSDKPDTRFDMLL D+

15 Sbjct: 243 SFMSQEDIMSLAEEMMAKVMRETKGEBELQLPLPRMTYDEAMNKYGSDDKPDTRFDMLLTDV 302

Query: 317 TEIVKEVDFKVFSEA---SVVKAIVVKDKADKYSRKNIDKLTEIAKQYGAKGLAWLKYA 372
 ++IVK+ +FKVFS A VVKAI VK A YSRK+ID L A YGAKGLAW+K

20 Sbjct: 303 SDIVKDTFEKVFSSAVANGGVVKAINVKGAGDYSRKDIDALGAFANYGAKGLAWVKVE 362

Query: 373 DNTISGPVAKFL-TAIEGRLTEALQLENNDLILFVADSLEVANETLGALRTRIAKELELI 431
 + + GP+AKF + +L EAL DL+LF AD EV +LGALR ++ KE LI

25 Sbjct: 363 ADGVKGPIAKFFDEEKQSKLIEALDAAEGDLLLFGADQFEVVAASLGALRLKLGKERGLI 422

Query: 432 DYSKFNFLWVVDWPMFEWSEEEGRYMSAHHPTLPTAETAHELEGLAKVRAVAYDIVLN 491
 D FNFLWV+DWP+ E EEGR+ +AHHPT+P E +E ++A AYD+VLN

30 Sbjct: 423 DEKLFNFLWVIDWPLLEHDPDEGRFYAAHHPFTMPVREDLELIETAPEDMKAQAYDLVLN 482

Query: 492 GYELGGGSLRINQKDTQERMFKALGFSAESAEQFGFLEAMDYGFPPHGGGLAIGLDRFV 551
 GYELGGGS+RI +KD QE+MF LGFS E A EQFGFLLEA +YG PPHGG+A+GLDR V

35 Sbjct: 483 GYELGGGSIRIFEKDIQEKMFALLGFSPEEAEBQFGFLEAFEGAPPHGGIALGLDRLV 542

Query: 552 MLLAGKDNIREVIAFPKNNKASDPMTQAPSLVSEQQLEELSLTVE 596
 MLLAG+ N+R+ IAFPK AS MT+AP VS+ QL+EL L+++

40 Sbjct: 543 MLLAGRTNLRDTIAFPKTASCLMTEAPGEVSDAQDELHLSIK 587

- 35 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3297> which encodes the amino acid sequence <SEQ ID 3298>. Analysis of this protein sequence reveals the following:

Possible site: 23
 >>> Seems to have an uncleavable N-term signal seq

40 ----- Final Results -----
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

- 45 An alignment of the GAS and GBS proteins is shown below.

Identities = 495/582 (85%), Positives = 538/582 (92%)

Query: 18 MKRSMYAGRVRSEHIGTSITLKGWVRRRDLGGLIFIDLDRDREGIMQLVINPEEVSASVM 77
 MKRSMYAGRVR EHIGT+ITLKGWV RRRDLGGLIFIDLDRDREG+MQLVINPEEVS+ VM

50 Sbjct: 18 MKRSMYAGRVRSEHIGTTITLKGWVRRRDLGGLIFIDLDRDREGVMQLVINPEEVSSDVM 77

Query: 78 ATAESLRSEBFVIEVSGVVTAREQANDNLPTGEVELKVQELSILNTSKTTPFEIKDGIEN 137
 ATAE LRSE+VIEV G V AR+QAND L TG VELKV L+ILNT+KTPFEIKD +E +

55 Sbjct: 78 ATAERLRSEYVIEVEGFVEARQQANDKLATGMVELKVSALTILNTAKTTPFEIKDDVEVS 137

Query: 138 DDTRMRYRYLDLRRPEMLENFKLRAKVTHSIRNYLDNLEFIDVETPMLTKSTPEGARDYL 197
 DDTR+RYRYLDLRRPEMLENFKLRAKVTHSIRNYLD+LEFIDVETPMLTKSTPEGARDYL

60 Sbjct: 138 DDTRLRYRYLDLRRPEMLENFKLRAKVTHSIRNYLDDLEFIDVETPMLTKSTPEGARDYL 197

Query: 198 VPSRVNQGHFYALPQSPQITKQLLMNAGFDRIYQIVKCFRDEDLRGDRQPEFTQVDLETS 257
 VPSRV+QGIFYALPQSPQITKQLLMNAGFDRIYQIVKCFRDEDLRGDRQPEFTQVDLETS

65 Sbjct: 198 VPSRVSQGHFYALPQSPQITKQLLMNAGFDRIYQIVKCFRDEDLRGDRQPEFTQVDLETS 257

Query: 258 FLSDQEIQDIVEGMIKVMKDTKGLEVSLEPFRMAYDDAMNNYGSDDKPDTRFDMLLQDLT 317
 FLS+QEIQDIVEGMIKVMK+TK ++V+LPFRM+YD AMN+YGSDDKPDTRF+MLLQDLT

-1193-

Sbjct: 258 FLSEQEIQDIVEGMIKVMKETKEIDVTLFFPRMSYDVAMNSYGSDDKPDTRFEMLLQDLT 317

Query: 318 EIVKEVDFKVFSEASVVKAIIVKDKADKYSRKNIDKLTETIAKQYGAAGLAWLKXADNTIS 377
VK DFKVFSEA VKAIVVK AD+YSRK+IDKLT AKQ+GAKGLAW+K D ++

5 Sbjct: 318 VTVKGNDFKVFSEAPAVKAIIVKGNADRYSRKDIDKLTETFAKQFGAKGLAWVKVTDGQLA 377

Query: 378 GPVAKFLTAIEGRLEALQLENNDLILFVADSLEVANETLGALRTIAKELIDYSEKFN 437
GPVAKFLTAIE L+ L+L NDL+LFVAD+LEVAN TLGALR RIAK+L++ID S+FN

10 Sbjct: 378 GPVAKFLTAIETELSSQLKLAENDLVLFVADTLEVANNLTGALRNRIAKDLDMIDQSQFN 437

Query: 438 FLWVVDWPMFEWSEEEGRYMSAHHPTLPTAETAHELEGLAKVRAVAYDIVLNGYELGG 497
FLWVVDWPMFEWSEEEGRYMSAHHPTLPT E+AHELEGLAKVRA+AYDIVLNGYELGG

15 Sbjct: 438 FLWVVDWPMFEWSEEEGRYMSAHHPTLPTESAHELEGLAKVRAIAYDIVLNGYELGG 497

Query: 498 GSLRINQKDTQERMFKALGFSAESAQEQFGLLEAMDYGFPPHGGLAIGLDRFVMLLAGK 557
GSLRINQK+ QERMFKALGF+A+ A +QFGLLEAMDYGFPPHGGLAIGLDRFVMLLAGK

20 Sbjct: 498 GSLRINQKEMQERMFKALGFSAESAQEQFGLLEAMDYGFPPHGGLAIGLDRFVMLLAGK 557

Query: 558 DNIREVIAFPKNNKASDPMTQAPSLVSEQQLEELSLTVESYE 599
DNIREVIAFPKNNKASDPMTQAPSLVSE QLEELSL +ES++

Sbjct: 558 DNIREVIAFPKNNKASDPMTQAPSLVSENNQLEELSLQIESHD 599

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

25 Example 1072

A DNA sequence (GBSx1146) was identified in *S. agalactiae* <SEQ ID 3299> which encodes the amino acid sequence <SEQ ID 3300>. Analysis of this protein sequence reveals the following:

Possible site: 54
>>> Seems to have no N-terminal signal sequence

30 INTEGRAL Likelihood = -8.44 Transmembrane 186 - 202 (182 - 205)
INTEGRAL Likelihood = -5.68 Transmembrane 88 - 104 (86 - 106)
INTEGRAL Likelihood = -3.40 Transmembrane 115 - 131 (112 - 132)
INTEGRAL Likelihood = -2.13 Transmembrane 141 - 157 (141 - 157)
INTEGRAL Likelihood = -0.96 Transmembrane 43 - 59 (43 - 59)

35 ----- Final Results -----
bacterial membrane --- Certainty=0.4376(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

40

The protein has homology with the following sequences in the GENPEPT database.

>GP:CAB12952 GB:Z99109 alternate gene name: yuxA-similar to
hypothetical proteins [Bacillus subtilis]
Identities = 104/275 (37%), Positives = 181/275 (65%), Gaps = 1/275 (0%)

45 Query: 39 EKISASLLYGILSSVAVNFFQPGHVYSSGATGLAQVISAVSKHWFSFEIPVALAFYAIN 98
+K+ ++ +L++ +N F P VY+SG TG+AQ++S+V + F I + +N
Sbjct: 7 KKLLIIVIIGALLNAAGLNLFLIPADVYASGFTGVAQLLSSVVDQYAPFYISTGTLLFLIN 66

50 Query: 99 IPLLILSWRKIGHKFTIFTITVTVSSIFIQLMPQITLTTPDPLINAIFGGLIMGAGVGFS 158
IP+ IL W K+G FT+++ ++V ++++F+ ++P+ +L+ D L+NA+FGG+I G+G +
Sbjct: 67 IPVGILGWLKVGKSFTVYSILSVALTTLFMGILPETSLSHDILLNAVFGGVISAVGIGLT 126

55 Query: 159 FKSRISSCGTDIISLTIRKKTGRDVGSISFIINGIILFAGLLFGWKYALYSMTIFVSS 218
K S+GG DI+++ + K + VG+ FI+NGII+L AGLL GW+ ALY++VT++V++
Sbjct: 127 LKYGASTGGLDIVAMVLAKWKDPVGTFFIILNGIILTAGLLQGWEKALYTLVTLYVIT 186

60 Query: 219 RVTDAIFTRQKKMQAMIVTSKPYCVIKRIHRDLHRGVTICINDAEGTYNHEKKAVLITILT 278
RV DAI T+ K+ AMIVT K + + I+ + RG+T + A+G + +E+K ++I ++T
Sbjct: 187 RVIDAIHTRHMKLTAMIVTKKADEIKEAIGKVMVRGITTV-PAKGAFNEQKEMMIIVIT 245

Query: 279 REEFSDFKYMLMLKADPKAFVSVSAENVHIIIGRFVDD 313

-1194-

R E D + ++ + DPKAF ++ + I G F D
 Sbjct: 246 RYELYDLEKIVKEVDPKAFNTNIVQTGTGIFGFFRKD 280

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3301> which encodes the amino acid
 5 sequence <SEQ ID 3302>. Analysis of this protein sequence reveals the following:

Possible site: 53
 >>> Seems to have no N-terminal signal sequence
 10 INTEGRAL Likelihood = -5.47 Transmembrane 87 - 103 (86 - 106)
 INTEGRAL Likelihood = -4.94 Transmembrane 185 - 201 (182 - 203)
 INTEGRAL Likelihood = -1.59 Transmembrane 114 - 130 (113 - 130)
 INTEGRAL Likelihood = -1.12 Transmembrane 42 - 58 (42 - 58)
 INTEGRAL Likelihood = -0.32 Transmembrane 140 - 156 (140 - 156)
 15 ----- Final Results -----
 bacterial membrane --- Certainty=0.3187(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

20 >GP:CAA66894 GB:X98238 orf2 [Lactobacillus sakei]
 Identities = 105/280 (37%), Positives = 180/280 (63%), Gaps = 7/280 (2%)
 Query: 37 AEKISASLLYGILSSIAVNFFFPQGHVYSSGATGLAQVFSAL-SHRLLG YDFPIAFAYL 95
 +++I +++YG L++++VN F P YSSG TG+AQ+ +AL SH LG +A ++
 25 Sbjct: 8 SKRIVIAMVYGFLAAVSVNLFILIPAKTYSSGVTGVAQLLTALVSH--LGGSL SVAALVFI 65
 Query: 96 INIPLLILAWYKIGHQFTIFTFTITVSMSSFFIQIMPQVT--LTTDPLINAI FGGLVMGMG 153
 +N+PLL+LAW+KI HQ+ IF+ + V S F++I+P + T+ A+FGG ++G+G
 30 Sbjct: 66 LNVPLLVLAWFKINHQAIFSI VAVFTSVIFLKIIPVPVQPILTERFAGALFGGALIGL 125
 Query: 154 IGTGLKSRISGGTDIVSLTLRKRTGKDVGSLSLMVNGAILAFAGILFGWQYALYSMVSI 213
 +G ++ S+GGTD++ + + TGK VG+++ ++NG I+ AGI FGW ALYS+V I
 Sbjct: 126 VGLCFRAGFSTGGTDVIVTLVGRITGKR VGAVNNVINGMIILAGIFFWGGAALYSIVEI 185
 35 Query: 214 FVSSRVTD AIFTKQKKMQATIVTSHPERVIHMIHKLHRGVTSINDAEGTYKHEQKAVLI 273
 FVSS + D I+T+Q+K+ TI T PE + + + +H G T + D G Y +++ +V++
 Sbjct: 186 FVSSLLMDYIYTQQQKVITVITFTKQPEALKKRMREFIH-GATEL-DGTGLYTNQETSVM 243
 Query: 274 TILTCEEYPEFKWMLKTDPAFVSAENVRIIGRFVEDD 313
 T+++ + K ++ DP AFV++ + + GRF ++
 40 Sbjct: 244 TVVSKYDLTALKLVVQDADPNAFVNIQSTMNLWGRFPESNE 283

An alignment of the GAS and GBS proteins is shown below.

Identities = 239/311 (76%), Positives = 274/311 (87%)
 45 Query: 4 RRTPLEKKVKYIISVWAKKFGLLHLTKSISREKYAEKISASLLYGILSSVAVNFFFPQGH 63
 ++T +KKVKY+IS AKK GLLH L+SISREKYAEKISASLLYGILSS+AVNFFFPQGH
 Sbjct: 3 KTTYKKVKYVISRGAKKVGLLHALRSTISREKYAEKISASLLYGILSSIAVNFFFPQGH 62
 50 Query: 64 VYSSGATGLAQVISAVSKHWFSFEIPVALAFYAINIPLLILSWRKIGHKFTIFTFTITVTV 123
 VYSSGATGLAQV SA+S ++ P+A AFY INIPLLIL+W KIGH+FTIFTFTITV++
 Sbjct: 63 VYSSGATGLAQVFSALSHRLLG YDFPIAFAYLINIPLLILAWYKIGHQFTIFTFTITVSM 122
 55 Query: 124 SSIFIQLMPQITLTTDPLINAI FGGLIMGAGVGFSSFKSRISGGTDIISLTIRKKTGRDV 183
 SS FIQ+MPQ+TLTTDPLINAI FGGL+MG G+G KSRISGGTDI+SLT+RK+TG+DV
 Sbjct: 123 SSFFFIQIMPQVTLTTDPLINAI FGGLVMGMGIGTGLKSRISGGTDIVSLTLRKRTGKDV 182
 Query: 184 GSISFIINGIILLFAGLLFGWKYALYSMTIFVSSRVTD AIFTKQKKMQAMIVTSKPYCV 243
 GS+S ++NG IL FAG+LFGW+YALYSMV+IFVSSRVTD AIFTKQKKMQA IVTS P V
 60 Sbjct: 183 GSLSLMVNGAILAFAGILFGWQYALYSMVSI FVSSRVTD AIFTKQKKMQATIVTSHPERV 242
 Query: 244 IKRIHRDLHRGVTCINDAEGTYNHEKKA VLITILTREEFSDFKYMLKADPKAFVSAEN 303
 I IH+ LHRGV T IND AEGTY HE+KAVLITILT EE+ +FK+LMLK DP+AFVSAEN
 Sbjct: 243 IHMIHKLHRGVTSINDAEGTYKHEQKAVLITILTCEEYPEFKWMLKTDPAFVSAEN 302

-1195-

Query: 304 VHIIGRFVDDD 314
 V IIGRFV+DD
 Sbjct: 303 VRIIGRFVEDD 313

5

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1073

A DNA sequence (GBSx1147) was identified in *S.agalactiae* <SEQ ID 3303> which encodes the amino acid sequence <SEQ ID 3304>. Analysis of this protein sequence reveals the following:

Possible site: 26
 >>> Seems to have a cleavable N-term signal seq.

INTEGRAL	Likelihood = -3.72	Transmembrane	156 - 172 (156 - 174)
INTEGRAL	Likelihood = -3.03	Transmembrane	112 - 128 (110 - 129)
INTEGRAL	Likelihood = -2.34	Transmembrane	80 - 96 (79 - 96)
INTEGRAL	Likelihood = -1.49	Transmembrane	60 - 76 (58 - 76)

----- Final Results -----

bacterial membrane	--- Certainty=0.2487(Affirmative) < succ>
bacterial outside	--- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm	--- Certainty=0.0000(Not Clear) < succ>

20

The protein has homology with the following sequences in the GENPEPT database.

>GP:BAB05397 GB:AF001512 unknown conserved protein [Bacillus halodurans]
 Identities = 113/278 (40%), Positives = 192/278 (68%), Gaps = 1/278 (0%)

Query: 7 KTKIKETILIAFGVALYTFGFKFNMNHLAEGGISGVTLLIHALFGVNPALSSLLLNIP 66
 + K K + I G A+++FG V FNM N+LAEGG +G+TLI++ +F +NPA+++L+LNIP
 Sbjct: 4 RLKWKNIIVFILLGSAIFSFGLVYFNMENNLAEKGFTGITLLIYFMFQINPAVTNLVLNIP 63

Query: 67 LFILGARILGKKSLLLTIYGTVLMSFFMWFWQQIP-VIVPLKNDMMLVAVAAGILAGTGS 125
 + ++G +ILG+ +L+ TI GTV +S F+ +Q+ + +PL +DM L A+ AG+ GTG
 Sbjct: 64 ILLIGWKILGRVTLLIYTIIGTVSVSVFLEMFQRWKFMDIPLHDDMTLAALFAGVFGTGL 123

Query: 126 GLVFRYGATTGCGADIIGRIVEEKGSGIKLGQTLLEFIDAIVLTSSLVYINLQQMLYTLVASF 185
 G+VFR+G TTGG DII ++ G +G+T+ DA+V+ SSL+Y+N ++ +YTL+A F
 Sbjct: 124 GIVFRFGGTTGGVDIIAKLGFRLGWSMGKTMFMFDDAVVIASSLIYLYNREAMYTL LAVF 183

Query: 186 VFSQVLINVENGGYTVRGMIIITKESESAATILHEINRGVTFLRGQAYSGREHDLVLYV 245
 + ++V+ ++ Y+ + II++ +E+ A TIL E+ RG T L+G+G+++G E ++LY
 Sbjct: 184 IAAKVIDFIQQTAYSAKAIFIIEHTEAIIADTILKEMERGATTLKKGSGFTGTEKEILYC 243

Query: 246 ALNPSEVRDVKEIMADLDPDAFISVINVDEVISSDFKI 283
 + +E+ +K ++ +DP AF++V +V +VI F +
 Sbjct: 244 VVGRNELIRLKS LVERIDPHAFVTVNDVQDVIGEGFTL 281

45

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3305> which encodes the amino acid sequence <SEQ ID 3306>. Analysis of this protein sequence reveals the following:

Possible site: 26
 >>> Seems to have a cleavable N-term signal seq.

INTEGRAL	Likelihood = -5.15	Transmembrane	112 - 128 (109 - 130)
INTEGRAL	Likelihood = -2.34	Transmembrane	156 - 172 (156 - 174)
INTEGRAL	Likelihood = -1.81	Transmembrane	178 - 194 (177 - 194)
INTEGRAL	Likelihood = -1.65	Transmembrane	80 - 96 (79 - 96)
INTEGRAL	Likelihood = -0.37	Transmembrane	60 - 76 (59 - 76)

----- Final Results -----

bacterial membrane	--- Certainty=0.3060(Affirmative) < succ>
bacterial outside	--- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm	--- Certainty=0.0000(Not Clear) < succ>

50

55

60

-1196-

The protein has homology with the following sequences in the databases:

>GP:BAB05397 GB:AP001512 unknown conserved protein [Bacillus halodurans]
Identities = 116/276 (42%), Positives = 182/276 (65%), Gaps = 1/276 (0%)

5 Query: 9 KLLKLFLLIALGVAIYTFGFVNFMNANALAEAGGVAGITLILHAHFGINPAYSSLLFNLPF 68
K + I LG AI++FG V FNM N LAEGG GITLIL+ F INPA ++L+ N+P+
Sbjct: 6 KWKNVIFILLGSAIFSPGLVYFNMENNLAEAGGFTGITLILYFMFQINPAVTNLVLNIPIL 65

10 Query: 69 ILGAKIFGKRSLALTIYGTVLMSAFIWMWQKVP-IELGLENDMMLVAVVAGLFSGIGSGI 127
++G KI G+ +L TI GTV +S F+ M+Q+ +++ L +DM L A+ AG+F G G GI
Sbjct: 66 LIGWKILGRVTLIYTIIGTVSVSVFLEMFQRWKFMDIPLHDDMTLALFAGVFGTGLGI 125

15 Query: 128 VFRYGATTGGTDIIGRIAEEKFGAKLGQTLLELDALVLTASLTIVDLKHMLYTLVASFVF 187
VFR+G TTGG DII ++ G +G+T+ + DA+V+ +SL Y++ + +YTL+A F+
Sbjct: 126 VFRFGGTGGVDIIAKLGFRYLGWSMGKTMFMFDDAVVIASSLIYLNRYEAMYTLAVFIA 185

20 Query: 188 SQMISVVQNGGYTIRGMIIITKHSEAAAQAILTEINRGVTYLGQGAYSGNDYNIMYVTL 247
+++I +Q Y+ + II++H+EA A IL E+ RG T LKG+G+++G + I+Y +
Sbjct: 186 AKVIDFIQQTAYSAKAAFIISEHTAIAIDTILKEMERGATTLKGKGSFTGTEKELYCVV 245

Query: 248 NPTEVREVKRILAGLDPDAFISIIDVDEVISSDFKI 283
E+ +K ++ +DP AF+++ DV +VI F +
Sbjct: 246 GRNELIRLKSLSVERIDPHAFVTVNDVQDVIGEGFTL 281

An alignment of the GAS and GBS proteins is shown below.

Identities = 206/286 (72%), Positives = 250/286 (87%)

30 Query: 5 DLKTKIKETILIAFGVALYTFGFVKFNMANHLAEGGISGVTLLIHALFGVNPALSSLLLN 64
D TK+ + LIA GVA+YTFGFV FNMN LAEGG++G+TLI+HA FG+NPA SSSL N
Sbjct: 5 DKLTKLLKLFLLIALGVAIYTFGFVNFMNANALAEAGGVAGITLILHAHFGINPAYSSLLFN 64

35 Query: 65 IPLFILGARILGKSLLLTIYGTVLMSFFMFWQQIPVTVPLKNDMMLVAVAAGILAGTG 124
+PLFILGA+I GK+SL LTIYGTVLMS F+W WQ++P+ + L+NDMMLVAV AG+ +G G
Sbjct: 65 LPLFILGAKIFGKRSLALTIYGTVLMSAFIWMWQKVP-IELGLENDMMLVAVVAGLFSGIG 124

40 Query: 125 SGLVFRYGATTGGADIIGRIVEEKSGIKLQQTLLFIDAIVLTSSLVYINLQQLYTLVAS 184
SG+VFRYGATTGG DIIGRI BEK G KLGQTL +DA+VLT+SL Y++L+ MLYTLVAS
Sbjct: 125 SGIVFRYGATTGGTDIIGRIAEEKFGAKLGQTLLELDALVLTASLTIVDLKHMLYTLVAS 184

45 Query: 185 FVFSQVLTNVENGGYTVRGMIIITKESASAAATILHEINRGVTFLRGQGAYSGREHDVLY 244
FVFSQ+++ V+NGGYT+RGMIIITK SE+AA IL EINRGVT+L+GQGAYSG ++++Y
Sbjct: 185 FVFSQMISVVQNGGYTIRGMIIITKHSEAAAQAILTEINRGVTYLGQGAYSGNDYNIMY 244

Query: 245 VALNPSEVRDVKEIMADLDPDAFISVINVDEVISSDFKIRRRNYDK 290
V LNP+EVR+VK I+A LDPDAFIS+I+VDEVISSDFKIRRRNYDK
Sbjct: 245 VTLNPTEVREVKRILAGLDPDAFISIIDVDEVISSDFKIRRRNYDK 290

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1074

A DNA sequence (GBSx1148) was identified in *S. agalactiae* <SEQ ID 3307> which encodes the amino acid sequence <SEQ ID 3308>. This protein is predicted to be BacB protein. Analysis of this protein sequence reveals the following:

55 Possible site: 60
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.4355(Affirmative) < succ>
60 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

-1197-

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database.

5 >GP:BAA11330 GB:D78257 BacB [Enterococcus faecalis]
 Identities = 27/88 (30%), Positives = 48/88 (53%), Gaps = 1/88 (1%)

Query: 1 MPSEKEILDALSKVYSEEVQADDYFRQAIFELASQLEKEGMN-SLLATKIDSLINQYVL 59
 M ++E+LD LSK Y++ I + + +FE A +L N + K+ ++ ++Y+
 10 Sbjct: 1 MDKQQELLDLLSKAYNDPKINEYEGLDKLFECARLTITNETNIGEVVCYKLSTINSEYLA 60

Query: 60 THQFDAPKSIFDLSRLVKTAKSHYKGTA 87
 H F+ PKSI +L + V + Y+G A
 Sbjct: 61 RHHFEMPKSIIEELQKFVTKEGQKYGWA 88

- 15 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3309> which encodes the amino acid sequence <SEQ ID 3310>. Analysis of this protein sequence reveals the following:

Possible site: 27
 >>> Seems to have no N-terminal signal sequence

20 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2712(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 25 An alignment of the GAS and GBS proteins is shown below.

Identities = 99/102 (97%), Positives = 102/102 (99%)

Query: 1 MPSEKEILDALSKVYSEEVQADDYFRQAIFELASQLEKEGMNSLLATKIDSLINQYVLT 60
 MPSEKEILDALSKVYSE+VIQADDYFRQAIFELASQLEKEGM+SLLATKIDSLINQY+LT
 30 Sbjct: 7 MPSEKEILDALSKVYSEQVIQADDYFRQAIFELASQLEKEGMSSLLATKIDSLINQYILT 66

Query: 61 HQFDAPKSIFDLSRLVKTAKSHYKGTAISAIMLGSFSLGGPK 102
 HQFDAPKSIFDLSRLVKTAKSHYKGTAISAIMLGSFSLGGPK
 35 Sbjct: 67 HQFDAPKSIFDLSRLVKTAKSHYKGTAISAIMLGSFSLGGPK 108

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 1075

- 40 A DNA sequence (GBSx1149) was identified in *S.agalactiae* <SEQ ID 3311> which encodes the amino acid sequence <SEQ ID 3312>. This protein is predicted to be ArgS (argS). Analysis of this protein sequence reveals the following:

Possible site: 57
 >>> Seems to have no N-terminal signal sequence

45 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2522(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

- 50 A related GBS nucleic acid sequence <SEQ ID 10271> which encodes amino acid sequence <SEQ ID 10272> was also identified.

The protein has homology with the following sequences in the GENPEPT database.

>GP:AAF86984 GB:AF282249 ArgS [Lactococcus lactis subsp. lactis]
 Identities = 377/566 (66%), Positives = 464/566 (81%), Gaps = 5/566 (0%)

55

-1198-

Query: 12 MDTKHLIASEIQKVVPD-MEQSTILSLLETPKNSSMGDLAFPAFSLAKTLRKAPQIIASD 70
MD K L++ + + I +++E PK+S +GDLAFPAF LAKTLRK+PQIIA +
Sbjct: 1 MDEKQLVSQLSAIDGVLGVEQIAAIEKPKSSDLGDLAFPAFQLAKTLRKSPQIIAGE 60

5 Query: 71 IAEQIKSDQFEKVEAVGPYVNFFLDKAAISSQVLKQVLSGSAAYATQNIGEGRNVAIDMS 130
IAE+I + FEKV AVGPYVNFFLDK A +S+V+++VL++G Y NIGEG NV IDMS
Sbjct: 61 IAEKIDTKGFEKVIAVGPYVNFFLDKNATASEVIREVLAEGEHYGDANIGEGGNVPIDMS 120

10 Query: 131 SPNIAKPFSIGHLRSTVIGDSLNIIFDKIGYHPVKINHLGDWQKQFGMLIVAYKKWGNNEE 190
+PNIAKPFSIGHLRSTVIGDS+A I++K+GY P+KINHLGDWQKQFG+LI AYKK+G+E
Sbjct: 121 APNIAKPFSIGHLRSTVIGDSIAKIYEKLGYPQIKINHLGDWQKQFGLLI+TAYKKYGDEA 180

15 Query: 191 AVRAHPIDECLKLVVRINAEAEETDPSVDEEAREWFRKLEANDPEATELWQWFRDESLEEF 250
+ A+PIDELLKLYV+INAE+ D VDEE R+WF K+E D EA +W+WF D SL+EF
Sbjct: 181 TITANPIDELLKLYVVKINAEAKEDSEVDEEGRQWFLKMEQGDEEALRIWKWFSVSLIEF 240

20 Query: 251 NRLYDQMNVTFDSYNGEAFYNDKMDDEVLELLESKNLLVESKGAQVVNLEKYGIEHPALIK 310
NR+Y ++ VTFD + GE+FY+DKMD ++E LE+KNLL ESKGA +V+LEKY + +PALIK
Sbjct: 241 NRIYKGLGVTFDFHMGESFYSDKMDAIVEDLENKNLLHESKGAIVDLEKYNL-NPALIK 299

25 Query: 311 KSDGATLYITRDLAAALYRKRTYDFAKSIYVVGNEQSAHFKQLKAVLKEMDYDWSDDMTH 370
K+DGATLYITRDLA A YRK+T++F KS+YVVG EQ+ HFKQLKAVLKE YDWSDDM H
Sbjct: 300 KTDGATLYITRDLATAAYRKKTFFNFVKSLEYVVGGEQTNHFKQLKAVLKEAGYDWSDDMVH 359

30 Query: 371 VPFGVLTKGAKLSTRKGNVILEPTVAEAINRAASQIEAKNPNLADKDKVAQAVGVGAI 430
VPFG+VT+GG K STRKG+V+ LE + EA++RA QIEAKNPNL +K++VA+ VGVGA+
Sbjct: 360 VPFGMVTQGGKKFSTRKGHVVKLEMALEAVDRAEKQIEAKNPNLENKEEVAKQVGVGAV 419

35 Query: 431 KFYDLKTDRTNGYDFDL+ MVSFEGETGPHYVQYAHARIQSILRKANFSPNSDNYSL--N 488
KFYDLKTDRTNGYDFDL+ MVSFEGETGPHYVQYAHARIQSILRKAN N DN SL +
Sbjct: 420 KFYDLKTDRTNGYDFDLDEMVSFEGETGPHYVQYAHARIQSILRKAN-RKVNIDNISLVVS 478

40 Query: 489 DVESWEIILQDFPRIIVRAADNFEPSIIAKFAINLAQCFNKYYAHTRILDEDAEISSR 548
D E+WEI+K +++FP I+ RAADN+EPSIIAK+AI+LAQ FNKYYAH RIL++DA++ R
Sbjct: 479 DAAWEIVKALKEFPNIVKRAADNFEPSIIAKYAIISLAQAFNKYYAHVRILEDDAQLDGR 538

Query: 549 LALCYATATVLKESLRLLGVDAPNEM 574
LAL AT+ VLKE+LRLLG V AP M
Sbjct: 539 LALISATSIVLKEALRLLGVAAPENM 564

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 3313> which encodes the amino acid sequence <SEQ ID 3314>. Analysis of this protein sequence reveals the following:

Possible site: 46
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.1734 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below.

Identities = 492/563 (87%), Positives = 526/563 (93%)

55 Query: 12 MDTKHLIASEIQKVVPDMEQSTILSLLETPKNSSMGDLAFPAFSLAKTLRKAPQIIASDI 71
MDTK LIASEI KVVP++EQ I +LLETPKNS MGDLAFFPAFSLAK LRKAPQ+IAS++
Sbjct: 1 MDTKTLIASEIAKVVELEQDAIFNLLETPKNSMDGDLAFPAFSLAKVLRKAPQMIASEL 60

60 Query: 72 AEQIKSDQFEKVEAVGPYVNFFLDKAAISSQVLKQVLSGSAAYATQNIGEGRNVAIDMSS 131
AEQI QFEKV AVGPY+NFFFLDKA ISSQVL+QV++ GS YA Q+ G+GRNVAIDMSS
Sbjct: 61 AEQIDESQFEKVAVGPYINFFFLDKAKISSQVLEQVITAGSDYAQQDEGQGRNVAIDMSS 120

65 Query: 132 PNIAKPFSIGHLRSTVIGDSLNIIFDKIGYHPVKINHLGDWQKQFGMLIVAYKKWGNNEE 191
PNIAKPFSIGHLRSTVIGDSLNI+IF K+GY PVKINHLGDWQKQFGMLIVAYKKWG+E A
Sbjct: 121 PNIAKPFSIGHLRSTVIGSLAHIFAKMGYKPVKINHLGDWQKQFGMLIVAYKKWGDEAA 180